informs
http://pubsonline.informs.org/journal/mksc/

MARKETING SCIENCE
Vol. 37, No. 4, July–August 2018, pp. 530–552
ISSN 0732-2399 (print), ISSN 1526-548X (online)

The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions

Raluca M. Ursu[a]

[a] Stern School of Business, New York University, New York, New York 10012

Contact: rursu@stern.nyu.edu, http://orcid.org/0000-0003-2239-1498 (RMU)

Received: April 3, 2016
Revised: May 6, 2017; July 24, 2017
Accepted: August 1, 2017
Published Online in Articles in Advance: June 29, 2018

https://doi.org/10.1287/mksc.2017.1072

Abstract. Online search intermediaries, such as Amazon or Expedia, use rankings (ordered lists) to present third-party sellers' products to consumers. These rankings decrease consumer search costs and increase the probability of a match with a seller, ultimately increasing consumer welfare. Constructing relevant rankings requires understanding their causal effect on consumer choices. However, this is challenging because rankings are endogenous: consumers pay more attention to highly ranked products, and intermediaries rank the most relevant products at the top. In this paper, I use the first data set with experimental variation in the ranking from a field experiment at Expedia to make three contributions. First, I identify the causal effect of rankings and show that they affect what consumers search, but conditional on search, do not affect purchases. Second, I quantify the effect of rankings using a sequential search model and find an average position effect of $1.92, which is lower than literature estimates obtained without experimental variation. I also use model predictions, data patterns, and a feature of the data set (opaque offers) to show rankings lower search costs, instead of affecting consumer expectations or utility. Finally, I show a utility-based ranking built on this model's estimates benefits consumers and the search intermediary.

History: K. Sudhir served as the editor-in-chief and Tat Chan served as associate editor for this article.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2017.1072.

Keywords: online consumer search • product rankings • sequential search model • endogeneity bias • online search intermediary

1. Introduction

The Internet has led to an explosion of product options facing consumers. Because evaluating these options is costly for consumers, in many industries, online search intermediaries have emerged. These intermediaries, such as Amazon or Expedia, return rankings (ordered lists) of third-party sellers' products to consumers in response to their queries, using algorithms that rank the most relevant products at the top. Intermediaries' rankings help consumers find a matching product more quickly, thereby decreasing search costs and increasing consumer welfare. Because of these benefits, consumers increasingly use search intermediaries in their search and purchase decisions. As consumers shift to using mobile devices to access Internet content, they become more impatient in their search, and hence the importance of intermediary rankings will further increase.

Constructing relevant rankings requires understanding their causal effect on consumer choices. For search intermediaries, measuring the causal impact of the ranking and separating it from the intrinsic quality of the product ranked allows them to place relevant products at the top of the ranking, rather than ones that were chosen more frequently merely because of their past rank. This helps consumers find better-matching products more quickly, which, in turn, benefits the search intermediary through increased conversions and a higher probability of repeat visits. For the sellers ranked, knowing the incremental value of a position allows them to measure the benefit of displaying their product on the search intermediary's platform, as well as to determine whether to bid for a sponsored slot (if sponsored slots are available). Finally, understanding the causal effect of rankings informs policy makers on whether consumer and search intermediary incentives are aligned, and thus on whether their intervention is required. For example, if rankings affect consumer search, but not their purchase decisions directly (as I will show is the case in this paper), the search intermediary is incentivized to rank those products the consumer will find relevant the highest, benefiting both. By contrast, if rankings affect consumer purchases directly, ranking less relevant products at the top might still lead to a purchase, because consumer and search intermediary incentives are not aligned.

Understanding the causal effect of rankings is challenging because rankings are endogenous: consumers pay more attention to highly ranked products, and intermediaries rank the most relevant products at the top. This produces a simultaneity problem in the estimation of the causal effect of rankings, which likely inflates the estimated effect, because products of higher quality are prioritized in the ranking. The literature has previously identified this problem and existing approaches correct for endogeneity using a number of methods, including the control function approach (De los Santos and Koulayev 2017), regression discontinuity design (Narayanan and Kalyanam 2015), simultaneous equation model (Ghose et al. 2014), and latent instrumental variables (Rutz et al. 2012). However, without experimental variation, measuring the causal effect of rankings on consumer choices is not conclusive.

In this paper, I employ the first data set with experimental variation in the ranking to study the causal effect of rankings on consumer choices. This data set comes from a field experiment run at Expedia, the world's largest online travel agent, where consumers looking for hotels were randomized into either (i) seeing Expedia's ranking, where hotels were ordered by relevance for consumers, or (ii) seeing a Random ranking, where hotels were listed in a random order. The data include 166,036 consumer queries for hotels along with their choices (search and purchases) over an eight-month period ending in June 2013. This corresponds to 4.5 million observations on the hotels displayed on Expedia.

Using this data set, I make three contributions. First, using the Random ranking, I show that the position of a product in the ranking has a causal effect on what consumers search, but conditional on search, do not affect purchase decisions. This finding suggests that intermediaries' rankings can influence consumer purchases only through their search decisions, emphasizing the importance of optimizing consumers' search. Under Expedia's ranking, however, rankings affect both search and conditional purchase decisions. To my knowledge, this difference empirically demonstrates for the first time the endogeneity bias in the ranking, namely, that the position effect is overestimated in the absence of experimental variation.

Second, to quantify the causal effect of rankings, I use a sequential search model following Weitzman (1979). The model will also be used to measure the welfare effects of a utility-based ranking. This model captures consumer behavior in an ordered environment and explains consumers' search and purchase decisions jointly. It has three components: consumer expected utility prior to search, expected utility from search, and search costs. To model the effect of position, I show that model predictions on the optimal search order, data patterns (specifically, the fact that rankings do not affect purchases conditional on search), and a feature of the data set called opaque offers, reject the effect of position on any of the three components or combinations of these, except the effect of position on search costs. Using this insight, I estimate the model using the Random ranking data and find an average position effect of \$1.92. As expected, because the Random ranking eliminates the endogeneity bias in the position effect, this estimate is typically lower than previous results in the literature without experimental variation in the ranking.

Third, I investigate the welfare effects of a utility-based ranking, obtained by reordering hotels based on their estimated expected utility prior to search. I find it leads to an increase in consumer welfare equaling on average \$30.24 (23.02% of the purchase price). A breakdown of this benefit shows that on average, 35% of the value is due to lower search costs, confirming the importance of modeling the effect of rankings on consumer search decisions. Finally, I show the utility-based ranking not only increases welfare but also benefits the search intermediary, with transactions increasing by at least 2.4%.

The rest of the paper is organized as follows. In Section 2 and 3, I review related work and I describe the institutional details of the online travel agent industry relevant for this paper, introduce the data, and provide reduced form evidence of the effect of position on consumer choices. In Section 4, I introduce the sequential search model, followed by a discussion of the estimation approach and identification. In Section 6, I present my results, and investigate the welfare effects of a utility-based ranking through a counterfactual. Section 7 concludes and provides a discussion of limitations and future research.

2. Related Literature

This paper relates to the literature on consumer search, in particular, to work examining the effect of rankings on consumer choices, which I emphasize in this section. Papers such as Chen and Yao (2016), De los Santos and Koulayev (2017), Koulayev (2014), and Ghose et al. (2012a, b, 2014) consider the effect of rankings on consumer online choices in the hotel industry and find a broad range of position effect estimates, as high as \$35 in De los Santos and Koulayev (2017). De los Santos and Koulayev (2017) and Ghose et al. (2014) also address the endogeneity problem of the ranking using a control function approach and a simultaneous equation model, respectively. By contrast, I find lower position effects by using the Random ranking to eliminate endogeneity.

Recently, several papers have used a sequential search model following Weitzman (1979) to quantify consumers' utility and search cost parameters. Kim et al. (2010) is one of the earliest papers to show how to use Weitzman (1979) to model consumer search decisions in an empirical setting, whereas Honka and Chintagunta (2016) extend it to also model purchase decisions. I use the same model to quantify the causal effect of rankings on both consumer search and purchase decisions. Furthermore, Chen and Yao (2016) and Ghose et al. (2012b) use a maximum likelihood approach that imposes restrictions on the parameters to be estimated to ensure consistency with optimal sequential search. My approach has the benefit of modeling the likelihood of observing consumer search and purchase decisions jointly, thereby capturing the complete set of parameter restrictions consistent with Weitzman's (1979) search rules. Finally, Kim et al. (2017) use the search rules in Weitzman (1979) in a probit model of sequential search. Their approach is an alternative to the one used in this paper that lowers the computational burden of estimation by providing semi-closed-form expressions for the probability of choice.

Exploring how to improve rankings, in particular, by using a utility-based method, is closely related to work in online recommender systems (see Ansari et al. 2000, and Ansari and Mela 2003) and has been the subject of several recent papers. Ghose et al. (2012a) was one of the earliest papers to use a utility-based ranking, which they show through lab experiments is superior to several baseline rankings. Furthermore, Ghose et al. (2014) show that a utility-based ranking outperforms other rankings in terms of revenues, whereas De los Santos and Koulayev (2017) show it can increase click-through rates almost twofold. Finally, Chen and Yao (2016) find that a utility-based ranking when consumers are informed about the ranking algorithm increases consumer utility by 2.4%. Similar to the literature, I find a utility-based ranking benefits both the consumer and the search intermediary. However, the utility-based ranking I consider is constructed using parameters estimated without the endogeneity bias of a curated ranking. Also, I decompose the change in welfare and find a large fraction of the improvement is due to lower search costs.

Understanding how rankings affect consumer search is a broader question that is also present in the online sponsored-search literature (Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011, Jerath et al. 2011, Yao and Mela 2011, Athey and Ellison 2011, Jeziorski and Segal 2015, Blake et al. 2015, Baye et al. 2016, Jeziorski and Moorthy 2018, Chan and Park 2015, Narayanan and Kalyanam 2015, Athey and Imbens 2015), as well as in the theoretical search literature (Hagiu and Jullien 2011, Berman and Katona 2013, De Corniere and Taylor 2014) and the emerging work on mobile search (Ghose et al. 2012c). In this literature, three papers are most relevant for my work. Baye et al. (2016) study search results at Google and Bing to measure the importance of name prominence and position on consumers' clicks. They also find that failing to account for the endogeneity in position inflates the position effect and minimizes the effect of name prominence. Narayanan and Kalyanam (2015) show how to use advertisers' quality scores in a regression discontinuity design framework to address the endogeneity of ad position, which I eliminate using the Random ranking. Most recently, Athey and Imbens (2015) focus on estimating heterogeneity in causal effects of rankings on clicks using data from an experiment demoting the best-matched search result to the third position. By contrast, I employ a data set in which rankings were fully randomized, and I also investigate the effect of rankings on purchases.

In sum, this paper draws on the vast literature emphasizing the importance of rankings in consumer choices, including work on the online hotel industry and work on online sponsored-search ads. It adds to this literature by measuring the causal effect of rankings, by employing a sequential search model to quantify this effect, and by investigating the welfare benefits of a utility-based ranking using data with experimental variation in the ranking.

3. Industry Background and Data Descriptive Analysis

3.1. The Online Travel Agent Industry

In this section, I describe the institutional details of the online travel agent industry that are relevant for this paper. In 2013, almost 80% of all bookings made online were made through online travel agents (OTAs), which had combined revenues of \$157 billion in the United States and \$278 billion worldwide.[1] In the United States, four OTAs, Expedia, Booking, Orbitz, and Travelocity, account for 95% of the bookings, with Expedia being the largest. OTAs revenue is derived under both the agency and the merchant model. Under the agency model, the OTAs receive a commission (e.g., ranging from 10% to 25% for hotels) from third-party sellers for a purchase. Under the merchant model, the OTA purchases the seller's product, which it marks up and sells to consumers.[2]

To compete for consumers, OTAs rank third-party sellers' products, such as flights, hotels, and rental cars. OTAs invest in constructing relevant rankings for consumers on the basis of machine learning techniques. In this section, I provide a general overview of one such ranking algorithm, and in Online Appendix B, I describe the technical details behind "learning to rank" algorithms. The position of a product in the ranking is a function of its historical purchases and clicks, its characteristics (e.g., price, quality), and its match with an individual consumer (e.g., consumer's past purchases). The algorithm learns a function of these components (i.e., a score) that best predicts the probability of a purchase or click and those products with a higher score are ranked at the top. In this case, sellers cannot directly affect their position in the ranking. In addition, OTAs reserve specific positions in the ranking for sponsored ads.[3] More precisely, a seller enters a pay-per-click auction for the sponsored ad slots by bidding to target consumer queries for particular destinations and travel dates.[4] The auction is adjudicated by evaluating the bids submitted by all sellers and their scores, and the winner is displayed in the same position on all result pages.

The institutional details of the OTA market motivate and affect my analysis. In particular, the algorithm used makes the position of a hotel endogenous. As a result, using observational data on consumer queries for hotels given the default ranking will not suffice in determining the causal effect of the ranking on consumer choices. Thus, I rely on estimates using the Random ranking that eliminate the endogeneity bias. Without properly separating the effect of position on choices from other hotel characteristics, the ability to improve the current ranking is limited.

3.2. Data

This data set comes from consumers' queries for hotels on Expedia.[5] It includes results from 166,036 queries along with consumers' clicks (interpreted as their search decisions) and purchases over an eight-month period between November 1, 2012, and June 30, 2013.[6] There are 4.5 million observations on 54,877 hotels located in 55 countries and 788 different destinations. The Expedia data are provided at the level of a search impression. A search impression is an ordered list of hotels and their characteristics seen by consumers in response to a query describing their trip.

The main feature of this data set is that in only two-thirds of the search impressions did consumers see Expedia's ranking, where hotels were ordered by relevance, while in one-third of the search impressions, hotels were ordered randomly (Random ranking). In Online Appendix B, I provide randomization checks to show that consumers were randomly assigned to the two rankings. I also show the position of the hotel was randomly generated under the Random ranking. To my knowledge, I am the first to use a data set with experimental variation in the ranking, that is, the Random ranking, to investigate the causal effect of rankings on consumer choices.

In addition to the Expedia data set that I use for the main analysis, I use a companion data set from the Wharton Customer Analytics Initiative (WCAI) on consumer queries for hotels on a popular OTA. This data set cannot be used to study the causal effect of rankings, because it does not include search impressions performed under a random ranking. However, it has information on some of the consumer groups that are excluded from the Expedia data set, as I describe below. I find the groups that are missing generally constitute a small fraction of search queries, allowing me to state that their absence will not affect the representativeness of the Expedia data.[7]

3.2.1. Search Process and Data Observed.

To describe the observables in this data set, in this section, I explain the three-step consumer search process on Expedia. First, the consumer begins her search query on Expedia by specifying details of her trip, such as the destination (city, country), the travel dates, and the number of travelers and rooms requested. In addition, based on her query, the number of days before the beginning of the trip is recorded (booking window). The data set includes information on all these variables. Second, in response to her query, the consumer gets a search impression of all of the hotels that match her request, distributed over multiple pages. From this search impression, I observe the first page of results displayed to consumers (the "list page"), which includes the hotel ID, its position in the ranking, and its characteristics (price, number of stars and reviews, location, a chain, and a promotion indicator).[8] On Expedia, consumers can sort or filter results; however, the data set only contains those search impressions where consumers made choices from the ranking displayed.[9] Third, after observing the list of hotels, the consumer can click on a particular one to observe more information. In this case, she navigates to a subpage reserved for that hotel (the "hotel page"), where she can see additional pictures, previous customers' reviews, and so on. Then she can either return to the previous screen to click on another hotel, leave the site without purchasing, or she can purchase. I observe all clicks and purchases consumers make. However, I do not observe the additional information consumers see on the hotel's page. Also, observations are provided at the search impression level, which means searches made by the same consumer cannot be linked.[10]

3.2.2. Data Description.

Table 1 provides summary statistics at the hotel and the search impression level. Hotels charge on average \$160 per night, have more than three stars, and a median review score of 4. Most hotels belong to a chain, and one quarter of them display a promotion. Location attractiveness is summarized by a score ranging from 0 to 7 designed by Expedia to measure how central a hotel is located, what amenities surround it, and so on, and the average hotel has a location score of 3.17. In a search impression, consumers see on average 27 hotels displayed. On average, consumers look for trips lasting two days that begin one month from their query. The median search impression is for a trip for one hotel room and two adults traveling with no children. The data set contains a total of 186,171 clicks, with 58,501 clicks under the Random ranking. There is at least one click per search impression, with 7% of search impressions

Table 1. Hotel and Search Impression Summary Statistics

	Observations	Mean	Median	SD	Min	Max
Hotel level						
Price	4,503,043	159.71	132.00	102.43	10	1,000
Stars	4,418,366	3.32	3.00	0.87	1	5
Review score	4,498,652	3.89	4.00	0.86	0	5
Chain	4,503,043	0.66	1.00	0.47	0	1
Location score	4,503,043	3.17	3.14	1.51	0	7
Promotion	4,503,043	0.25	0.00	0.44	0	1
Search impression level						
Number of hotels displayed	166,036	27.12	31.00	8.10	5	38
Trip length (days)	166,036	2.42	2.00	1.98	1	40
Booking window (days)	166,036	39.26	18.00	53.89	0	498
Saturday night (percent)	166,036	0.50	1.00	0.50	0	1
Adults	166,036	2.00	2.00	0.90	1	9
Children	166,036	0.39	0.00	0.79	0	9
Rooms	166,036	1.12	1.00	0.44	1	8
Total clicks	166,036	1.12	1.00	0.61	1	25
Two or more clicks (percent)	166,036	0.07	0.00	0.25	0	1
Total transactions	166,036	0.66	1.00	0.48	0	1
Random ranking (percent)	166,036	0.31	0.00	0.46	0	1

including two or more clicks, suggesting that consumer search costs are large. Two-thirds of all search impressions end in a transaction for a total of 108,903 transactions overall and 3,930 under the Random ranking. Clicked (purchased) hotels are on average \$13 (\$22) cheaper than displayed hotels, have a higher number of stars and reviews, and run more frequent promotions (for details, see Table 3 in Online Appendix B). Approximately 7,500 search impressions have historical information on the average star rating and price of hotels previously purchased by a consumer (not summarized here).

The data are anonymized, so identifying the destination to which a consumer wishes to travel is not possible. However, evidence suggests the largest country is the United States.[11] In the data, 80% of queries to this country are made by consumers that live there, suggesting it has a large territory with a large fraction of domestic travel. This evidence is also consistent with the fact that 73% of Expedia's traffic comes from U.S. visitors.[12]

3.2.3. Pros and Cons of Using the Expedia Data Set. The main advantage of using this Expedia data set comes from the Random ranking, which provides the unique opportunity to study the causal effect of rankings on consumer choices. However, two features of the data limit the analysis in this study: (i) search impressions included in the data contain at least one click, and (ii) search impressions leading to a transaction were oversampled.[13] These features have three implications for the analysis. First, the fact that only search impressions with at least one click are observed means consumers who did not click are missing from the sample, and hence the results may not generalize to all consumers who saw the Random ranking. Second, the fact that only search impressions with at least one click are observed for the two rankings means they may correspond to distinct fractions of the consumer populations who saw the respective rankings, and hence a quantitative comparison of consumers' choices under the two rankings may be problematic. Even though I show that consumers were randomly assigned to seeing either ranking and exhibit similar observable characteristics in the sample (see Online Appendix B), a quantitative comparison of the two rankings is still difficult because consumers may differ on unobservables (e.g., consumers who click under the Random ranking may have lower search costs than those who click under Expedia's ranking). Third, the fact that converting search impressions were oversampled means that the total number of purchases cannot be compared across the two rankings.[14] In what follows, I show how I was able to overcome these limitations to study the causal effect of rankings.

For the reduced form part of the paper (Section 3.3), I use the Random ranking to compare the fraction of clicks and purchases across positions. The sampling done by Expedia may affect the absolute number of clicks and purchases by position, but the relative effect of position on choices is unchanged. Because this relative effect recovers the causal effect of rankings, the sampling done does not affect the results. If search impressions contain at least one click, and converting search impressions were oversampled, another concern is that consumers have identified a hotel to purchase on previous visits to the site, which would minimize the observed effect of rankings on their choices. However, this story is contradicted by the fact that consumers click more often on higher-ranked hotels under the Random ranking, which are unlikely to display their previously identified hotel (see Section 3.3). For this part of the paper, I use the Expedia ranking only to point out qualitatively the direction of the bias that would arise in the absence of experimental variation, without quantifying differences between the two rankings.

For the sequential search model, I use the Random ranking to quantify the effect of rankings on consumer choices. In this case, removing search impressions without clicks may translate into an underestimation of search costs. However, the oversampling of transactions does not affect utility and search cost parameter estimates: utility parameters can be consistently estimated because I observe both consumers who purchase and those who do not, whereas search costs are not identified from purchase decisions and are thus not affected by the sampling done (see Section 5.2 for details).

For the counterfactual, I compare the utility-based ranking built on the model's estimates with simulated choices under Expedia's ranking (instead of actual choices). Simulated choices are computed using the utility and search cost parameters estimated using the Random ranking and Weitzman's (1979) search rules. By doing so, I avoid the direct comparison of consumer choices under the two rankings, which may be biased because their underlying parameters may be different because of the sampling done.

Despite its limitations, I conclude the causal effect of rankings on choices can be studied with this data set that contains experimental variation in the ranking.

3.3. Reduced Form Evidence

In this section, I present reduced form results of the causal effect of rankings. More precisely, I look at the effect of position on consumer search and purchase decisions under the Random ranking, which does not suffer from endogeneity bias. My results can be found in Figure 1 and Table 2.

In the left panel of Figure 1, I illustrate the click-through rate of a position, that is, the fraction of times a position was clicked out of all of the times it was displayed.[15] I find that the higher hotels are ranked, the more clicks they receive, which happens despite the fact that under the Random ranking, the quality of hotels is unrelated to their position in the ranking. This proves rankings have a causal effect on consumer search.[16] In the right panel of Figure 1, I plot the conversion rate of a position, that is, the percent of clicks that end in a purchase. I find that the conversion rate of the Random ranking is approximately constant across positions. Thus, rankings have no causal effect on consumers' purchases conditional on search. Table 2 shows the same pattern as in Figure 1 holds in a regression controlling for hotel, destination, and query characteristics. In particular, hotels ranked lower (higher position) have a lower probability of being clicked (columns 1–3) and purchased (columns 4–5), whereas conditional on a click, the position of the hotel has no effect on purchases (columns 6–8). This result holds both when considering the linear effect of position (columns 1, 2, 4, 6, and 7), the nonlinear fixed effect (columns 3, 5, and 8), and when using a logit model instead of a linear probability model (see Table 4 in Online Appendix B). Note that unconditional on a click, higher-ranked hotels lead to more purchases (as results in columns 4 and 5 in Table 2, as well as those in Figure 3 in Online Appendix B show). However, the fact that conditional on a click, consumers who clicked at the top and those who clicked lower in the ranking are equally likely to purchase means intermediaries' rankings can influence consumer purchases only though their search decisions.

To put these results into context, in Figure 2, I plot the click-through rate and the conversion rate of Expedia's ranking. As described in Section 3.2.3, I cannot quantitatively compare the two rankings.[17] However, I can point out qualitatively what the pattern in click-through rate and conversion rate is in the absence of experimental variation in the ranking. In contrast to the results for the Random ranking, Figure 2 shows that under Expedia's ranking, higher-ranked hotels lead to

Figure 1. (Color online) The Effect of Position on Click-Through Rate and Conversion Rate: Random Ranking

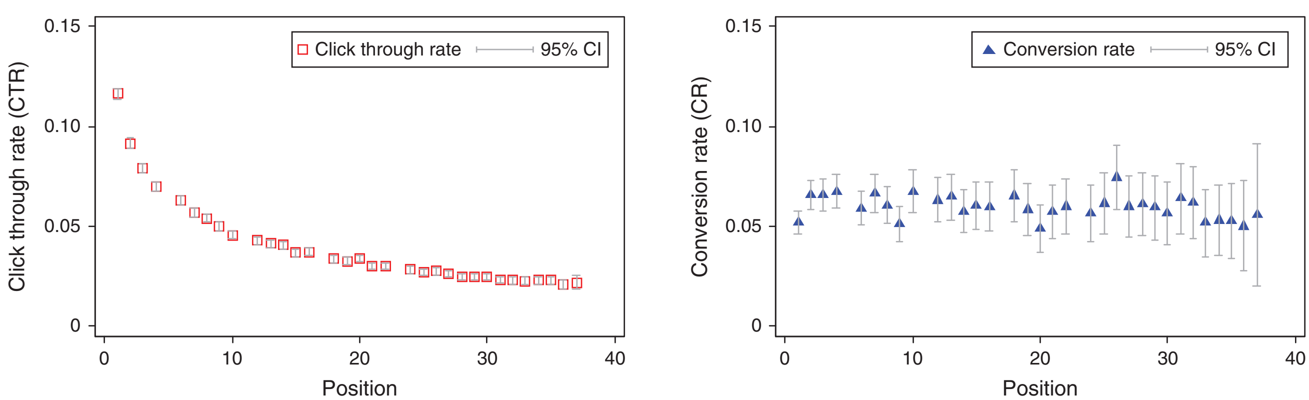


Table 2. Estimates of Click, Transaction, and Transaction Conditional on Click (OLS): Random Ranking

	Click			Transaction		Transaction conditional on click		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Position effect								
Position	−0.0019***	−0.0019***		−0.0001***		−0.0001	−0.0000	
	(0.0000)	(0.0000)		(0.0000)		(0.0001)	(0.0001)	
Position 1			0.0918***		0.0051***			0.0061
			(0.0023)		(0.0005)			(0.0120)
Position 2			0.0677***		0.0051***			0.0204
			(0.0022)		(0.0005)			(0.0122)
Position 3			0.0549***		0.0043***			0.0206
			(0.0021)		(0.0005)			(0.0124)
Position 4			0.0462***		0.0038***			0.0220
			(0.0021)		(0.0005)			(0.0124)
Position 6			0.0390***		0.0028***			0.0126
			(0.0021)		(0.0005)			(0.0125)
Position 7			0.0335***		0.0029***			0.0212
			(0.0021)		(0.0005)			(0.0126)
Position 8			0.0304***		0.0024***			0.0140
			(0.0020)		(0.0005)			(0.0126)
Position 9			0.0266***		0.0017***			0.0048
			(0.0020)		(0.0005)			(0.0126)
Position 10			0.0221***		0.0022***			0.0220
			(0.0020)		(0.0005)			(0.0129)
Hotel characteristics								
Price		−0.0001***	−0.0001***	−0.0000***	−0.0000***		−0.0001***	−0.0001***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)
Stars		0.0161***	0.0161***	0.0011***	0.0011***		0.0009	0.0009
		(0.0003)	(0.0003)	(0.0001)	(0.0001)		(0.0017)	(0.0017)
Review score		0.0012***	0.0013***	0.0002***	0.0002***		0.0061***	0.0061***
		(0.0002)	(0.0002)	(0.0000)	(0.0000)		(0.0011)	(0.0011)
Chain		0.0022***	0.0023***	0.0003*	0.0003*		0.0047	0.0045
		(0.0005)	(0.0005)	(0.0001)	(0.0001)		(0.0024)	(0.0024)
Location score		0.0044***	0.0044***	0.0005***	0.0005***		0.0072***	0.0072***
		(0.0002)	(0.0002)	(0.0000)	(0.0000)		(0.0008)	(0.0008)
Promotion		0.0116***	0.0114***	0.0012***	0.0012***		0.0104***	0.0104***
		(0.0005)	(0.0005)	(0.0001)	(0.0001)		(0.0024)	(0.0024)
Query characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Adjusted R^2	0.010	0.015	0.018	0.003	0.003	−0.000	0.033	0.033
Observations	1,245,455	1,220,917	1,220,917	1,220,917	1,220,917	54,614	53,921	53,921

Notes. Standard errors clustered at the search impression level (standard errors are in parentheses). Regressions in columns (3), (5), and (8) include position FE for positions 1 through 37, and coefficients should be interpreted with respect to the omitted category: the last position. For visibility, I only display the coefficients for positions 1–10. I restrict the sample to search impressions with opaque offers, which means no hotel was displayed in positions 5, 11, 17, and 23 in the ranking observed by consumers. Query characteristics included in all regressions: trip length, booking window, number of adults and children traveling, number of rooms, and an indicator for a Saturday-night stay.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

both more clicks and more purchases conditional on a click; hence, the effect of position is increased (Tables 5 and 6 in Online Appendix B show the regression results mirroring the pattern shown in Figure 2). This difference empirically demonstrates, for the first time to my knowledge, the endogeneity bias in the ranking, namely, that the position effect is overestimated in the absence of experimental variation in the ranking.

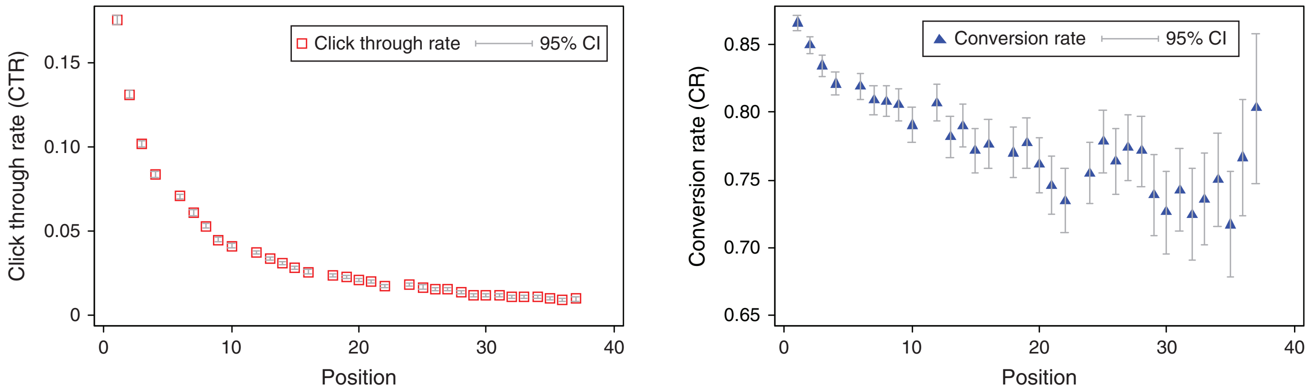
4. Model

In this section, I develop a sequential search model following Weitzman (1979), which will be used in the following sections to quantify the effect of rankings on consumer search and purchase decisions and to measure the welfare effects of a utility-based ranking.

4.1. Model Components

Consider the decision of a consumer who is searching for a product on a search site, such as Expedia.[18] This site provides a list of J ordered products from which she can choose. The list page contains some information about the displayed products, whereas each product's page contains additional information accessible by clicking the product on the list page. For example,

Figure 2. (Color online) The Effect of Position on Click-Through Rate and Conversion Rate: Expedia Ranking



Note. The scale of the y-axis in the two figures is different.

in the current application, if the consumer searches for a hotel on Expedia, the list-page information contains the hotel name, number of stars, price, promotion, and average reviews of the hotel; whereas the hotel page contains additional information, including pictures, amenities, and location of the hotel on the map. The consumer decides which alternatives to search (by clicking) and whether to purchase.

To capture this setting, let v_j denote the consumer valuation of the list-page information for product j and let ϵ_j denote her valuation of the product-page information.[19] The consumer knows v_j without clicking, but must search to reveal ϵ_j.[20] Thus, in what follows, I will refer to v_j as the expected utility prior to search, and to ϵ_j as the expected utility from search. After search, the realized utility of j is given by $u_j = v_j + \epsilon_j$. In most settings, ϵ_j is unobserved by the researcher and is modeled as an independent draw. Following Kim et al. (2010, 2017) and Chen and Yao (2016), I assume ϵ_j follows a normal distribution with mean zero and product uncertainty (standard deviation) σ_j. Consumers search to reveal ϵ_j by paying a search cost to click the product on the list page. Search costs may represent consumers' opportunity cost of time spent determining their match with a product, that is time spent on the product page, or search costs might capture consumers' mental cost from processing product-page information. At the same time, as demonstrated in Section 3, lower positions (closer to the top) are more accessible. To capture both of these effects, I model search costs $c_j(p_j)$ as a function of the position p_j of the product in the list, with $c_j(p_j) > 0$ and $c_j'(p_j) > 0$ (Ghose et al. 2012b, Chen and Yao 2016). The mean search cost component captures the opportunity cost of time or the mental cost, whereas the position effect captures the fact that lower-ranked products are more accessible (for more details, see the empirical specification of the model in Section 4.4). In Section 4.3, I provide evidence to support the assumption that the position of the product on the list only affects search costs, without affecting v_j or ϵ_j. Finally, the consumer can decide to forgo the purchase, in which case she may choose the (outside) option $j = 0$ available at no cost with expected utility prior to search equal to a known constant, v_0.

In sum, the model has three components:

- *Expected utility prior to search* v_j, for $j \in \{0, 1, \ldots, J\}$.
- *Expected utility from search* $\epsilon_j \sim N(0, \sigma_j^2)$, with $\sigma_j > 0$, for $j \in \{0, 1, \ldots, J\}$.
- *Search cost* $c_j(p_j)$, for $j \in \{1, \ldots, J\}$.

4.2. Optimal Search

To study consumers' optimal search strategy, consider the following scenario. Suppose the consumer has searched a number of options (as well as the outside option, which is always searched). She must then decide whether to continue searching, and if so, which option to search. If she decides to stop searching, she must determine whether to purchase, and if so, which of the searched options to choose. To model this behavior, I rely on Weitzman (1979), who characterizes consumers' optimal search strategy using the following search rules.

1. *Selection rule*. If a search is to be made, the option with the highest reservation utility should be searched next.

The reservation utility z_j of option j is defined as the level of utility that would make the consumer indifferent between searching it or not. It is computed by equating the marginal gain from searching j with the marginal cost from

$$c_j = \int_{z_j}^{\infty} (u_j - z_j) f(u_j)\, du_j, \tag{1}$$

where $f(\cdot)$ is the probability density function of u_j.

2. *Stopping rule*. Search should terminate when the maximum utility observed exceeds the reservation utility of any unsearched option.

3. *Choice rule*. Once the consumer stops searching, she will choose the option with the highest utility among those searched.

Following Kim et al. (2010), Equation (6) can be rewritten as

$$\begin{aligned} c_j &= \int_{z_j}^{\infty} (u_j - z_j) f(u_j)\,du_j \\ &= (1-F(z_j))\left[\int_{z_j}^{\infty} (u_j - z_j)\frac{f(u_j)}{1-F(z_j)}\,du_j\right] \\ &= \left(1-\Phi\left(\frac{z_j - v_j}{\sigma_j}\right)\right)\left(v_j - z_j + \sigma_j \frac{\phi((z_j - v_j)/\sigma_j)}{1-\Phi((z_j - v_j)/\sigma_j)}\right). \end{aligned} \quad (2)$$

Dividing by σ_j and defining $x_j = c_j/\sigma_j$ and $m_j = (z_j - v_j)/\sigma_j$, the equation becomes

$$\begin{aligned} x_j &= (1-\Phi(m_j))(\lambda(m_j) - m_j) \\ &= B(m_j), \end{aligned} \quad (3)$$

where $\lambda(\cdot) = \phi(\cdot)/(1-\Phi(\cdot))$ is the hazard rate. When $1-\Phi(m_j) > 0$, a unique pair of x_j and m_j solves $x_j = B(m_j)$ and the inverse $m_j = B^{-1}(x_j)$ exists. Thus, given search costs and product uncertainty, one can invert function B to obtain m_j, and thereby compute the reservation utility of product j from

$$z_j = v_j + \sigma_j m_j. \quad (4)$$

Obtaining an expression for the reservation utility is essential because Weitzman's (1979) optimal search rules describe relationships between the reservation utility and the realized utility.

4.3. Modeling the Effect of Position

In this section, I explore all possible ways in which the position of a product in the list can affect consumer decisions in the model presented above. In this model, the position p_j of a product j can affect consumer decisions by affecting any of its three components, v_j, ϵ_j, or c_j separately or any combination of these components. Thus, a total of seven possibilities exist, summarized in Table 3 (dropping subscripts). Note that position can affect ϵ_j either by affecting its distribution $N(0,\sigma^2)$ prior to search, or by affecting its realization after search. However, it is sufficient to consider only the former effect, which is what I do in this section, since an effect of position on the realized ϵ_j is equivalent to an effect of position on realized utility and is thus studied with the same tools as those used in Cases 4, 5, or 7.[21] In what follows, I show that model predictions on the optimal search order, data patterns (specifically, the fact that rankings do not affect purchases conditional on search), and a feature of the data set called opaque offers, reject the effect of position on any of the three components or combinations of these, except the effect of position on search costs. It should be emphasized that these results hold under the specific assumptions of the model presented in Section 4.1. Most importantly, the assumption that ϵ_j follows a normal distribution allows me to use functional relationships between model components and reservation utilities to examine the effect of position in this model.

Table 3. Possible Ways of Modeling the Effect of Position

Case	1	2	3	4	5	6	7
Position affects	v	ϵ	c	v, ϵ	v, c	ϵ, c	v, ϵ, c

Case 1: Position Affects Only v. Suppose position affects the expected utility prior to search, v, without affecting ϵ or c. In this case, the realized utility is also a function of position, because $u = v + \epsilon$. According to Weitzman's (1979) choice rule, once the consumer stops searching, she will purchase the alternative with the largest realized utility (including the outside option). If the realized utility is a function of position, then the sequential search model predicts that the position will affect the purchase decision, conditional on search.

Prediction 1 (Case 1). *If position affects the expected utility prior to search, then conditional on search, position affects purchases.*

However, the evidence presented in Section 3.3 contradicts this prediction. More precisely, the regression results for the Random ranking presented in Table 2 and the conversion rate pictured in Figure 1, show position has no causal effect on purchases conditional on a click. Therefore, I conclude that position cannot affect only v, which rules out Case 1.

Case 2: Position Affects Only ϵ. Suppose position affects ϵ, without affecting v or c. This may occur if consumer beliefs about what product qualities they will reveal upon search are different for each positions (for example, higher at the top of the ranking if consumers expect Expedia to rank first the most relevant products). Thus, even if position does not affect v or purchases conditional on search, consumer choices before search may be influenced by position if this affects ϵ.[22] Note that position might affect ϵ even if Expedia changes the ranking to the Random ranking, but consumers are not informed of the change, since consumer beliefs in the ranking algorithm are unaltered in this case. In addition, the effect of position on ϵ can explain the evidence provided in Section 3.3, showing that higher ranked options have a higher click-through rate, but conditional on click, have a similar conversion rate to lower ranked options. Nevertheless, in what follows, I show that model predictions on consumer search order rule out the effect of position on ϵ.

If position affects ϵ, because $\epsilon \sim N(0,\sigma^2)$, the only way in which position can affect the consumer's expected utility from search is if it affects σ, the product uncertainty.[23] If product uncertainty depends on position, $\sigma(p)$, then it is reasonable to assume that

$\sigma'(p) > 0$,[24] that is, uncertainty increases in position, so the consumer is more certain about the expected utility from search of higher-ranked alternatives (those closer to the top), conditional on list-page characteristics observed (v). This assumption is reasonable given that the ranking algorithm favors hotels with higher purchase rates, thereby favoring hotels that previous consumers deemed relevant after taking product-page, as well as list-page information, into account.[25] Thus, the position of a higher-ranked product may be a more accurate predictor of product-page information given list-page information, as captured by the assumption $\sigma'(p) > 0$.

A well-established result of sequential search models is that, all else equal, searching first those products with rewards drawn from distributions that are riskier is optimal (see Weitzman 1979), because the reservation utility of an option (which dictates search order) is unaffected by the lower tail of the utility distribution, as can be seen in the optimal stopping Equation (6). By contrast, a product with rewards drawn from a distribution that is more spread out at the upper tail, has a higher reservation utility, because searching this option may mean terminating search earlier. Under the assumption that $\sigma'(p) > 0$, I can formulate the following prediction for Case 2.

Prediction 2 (Case 2). *If position affects only the expected utility from search, and product uncertainty is increasing in position, the selection rule of the sequential search model dictates that the reservation utility is increasing in position.*

In what follows, I will first show that the current model produces this prediction, and then proceed to test it. As derived in Equation (4), the reservation utility is defined by $z = v + \sigma m$ where $m = B^{-1}(x)$ and $x = c/\sigma$. Differentiate this equation by p on both sides to obtain

$$z' = \sigma' m + \sigma m', \tag{5}$$

where $m' = (\partial m/\partial x)(\partial x/\partial p)$. To simplify Equation (5), note that $x = B(m) = (1 - \Phi(m))(\lambda(m) - m) = \phi(m) - m(1 - \Phi(m))$ because $\lambda(m) = \phi(m)/(1 - \Phi(m))$. Then, using the fact that for the standard normal density $\phi'(m) = -m\phi$, I obtain

$$\frac{\partial x}{\partial m} = -(1 - \Phi(m)) < 0.$$

Using the relation $m = B^{-1}(x)$ and the definition of the derivative of the inverse of a function, it follows that

$$\frac{\partial m}{\partial x} = \frac{\partial B^{-1}(x)}{\partial x} = -\frac{1}{1 - \Phi(B^{-1}(x))} = -\frac{1}{1 - \Phi(m)} < 0. \tag{6}$$

Because $x = c/\sigma$,

$$\frac{\partial x}{\partial p} = \frac{-c\sigma'}{\sigma^2} = \frac{-x\sigma'}{\sigma}, \tag{7}$$

which is negative because search costs are positive and $\sigma' > 0$ by assumption. Using Equations (6) and (7) to express m', Equation (5) can be rewritten as

$$\begin{aligned} z' &= \sigma'\left[m + \frac{x}{1 - \Phi(m)}\right] \\ &= \sigma'[m + (\lambda(m) - m)] \\ &= \sigma'\lambda(m), \end{aligned}$$

where the second equality follows from $x = (1 - \Phi(m))(\lambda(m) - m)$. Because $\lambda(m) > 0$, then $z' > 0$ under the assumption that $\sigma' > 0$. Thus, if position only affects ϵ and $\sigma' > 0$, then the model predicts that, all else equal, the reservation utility z of an alternative is higher for alternatives displayed lower in the ranking (higher p), which is what Prediction 2 describes.

To test this prediction, one has to observe not only consumer clicks and the position of the products clicked but also the order in which clicks occur. The Expedia data set does not include information on click order. Fortunately, the WCAI companion data set contains this information. Thus, I will use the WCAI data set to check whether consumers with at least two clicks, first click lower in the ranking, all else equal.

Table 4 tests the prediction of Case 2 in the data by looking at the effect of position on click order, conditional on observed hotel characteristics. More precisely, the first three columns show results from a regression of an indicator for the first click on the position of a hotel in the list, the fourth column considers a logit model with the same dependent variable, and the last column uses a rank-ordered logit model to model the effect of position on click order. In all cases, I find that position has a negative effect on click order, that is consumers click first on alternatives displayed closer to the top (lower position), conditional on hotel characteristics observed. Thus, these results contradict the predictions of a model in which position affects only the expected utility from search, ϵ, thereby ruling out Case 2.

Case 3: Position Affects c. Suppose position only affects search costs, so $c(p)$, without affecting v or ϵ. Then, because the choice rule is unaffected by search costs, the test for Prediction 1 does not rule out Case 3. Also, if position only affects search costs, then $z' = -1/(1 - \Phi(m)) < 0$, so that the test for Prediction 2 does not rule out Case 3. To determine the effect of position on search costs, an ideal test would involve a search cost shifter, such as an ad inserted into the ranking, satisfying three criteria: (i) keeping the expected utility prior to search, v, constant; (ii) keeping the expected utility from search, ϵ, constant; (iii) providing an exogenous shifter unrelated to consumer characteristics.

A feature of the Expedia data set provides such a search cost shifter. More precisely, Expedia reserves

Table 4. Effect of Position on Click Order (WCAI Manhattan)

	OLS DV: First click (first two clicks)	OLS DV: First click	OLS DV: First click (first page)	Logit DV: First click	Rank-ord. logit DV: Click order
	(1)	(2)	(3)	(4)	(5)
Position	−0.0112***	−0.0149***	−0.0129***	−0.0779***	−0.1074***
	(0.0015)	(0.0011)	(0.0014)	(0.0058)	(0.0057)
Hotel characteristics	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.027	0.058	0.040		
Log-likelihood				−2,083	−2,658
Observations	2,192	3,575	2,211	3,575	5,473

Notes. Columns (1) through (4) show regressions of an indicator for the first click on position and hotel characteristics. Column (5) shows results from a rank-ordered logistic regression of click order on position and hotel characteristics. For this analysis, I restrict attention to search impressions with at least two clicks. Column (1) further restricts attention to the first two clicks made, while column (3) considers only clicks made on the first page of results. All regressions include controls for the following observable hotel characteristics: stars, review score, price, and promotion and chain indicators. Standard errors are in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

certain positions in the ranking (such as 5, 11, 17, and 23) to display opaque offers.[26] Expedia displays opaque offers (i.e., "Expedia Unpublished Rate Hotels") on its website using Hotwire's inventory (a company owned by Expedia). Hotwire sells travel inventory of branded hotels at discounted prices (typically 30%–60% discount), and in 2012, opaque offers accounted for 6% of the sales of such hotels. The consumer has some information on the hotels sold as opaque offers (e.g., the star rating of the hotel, its general location, and the discounted price), but does not know the name of the hotel before making a purchase (hence the name "opaque"). This practice allows hotels to prevent cannibalization of their regular priced rooms.

Importantly for this study, opaque offers occupy a position in the ranking, but are not one of the ranked hotels. This means opaque offers satisfy criterion (i), because they do not affect the characteristics of the hotels displayed. In addition, when an opaque offer is displayed, consumers' expectations from searching one of the ranked hotels should not change, because their order in the search impression has not changed (the nth best hotel according to Expedia's algorithm is still the nth displayed hotel, although it may not appear in position n). Therefore, opaque offers also satisfy criterion (ii). Finally, the opaque offer demotes a hotel to a lower position, increasing its search cost. The variation in the position of hotels due to opaque offers is unrelated to individual consumer queries (exogenous), because offers are displayed based on availability in the data (thereby satisfying criterion (iii)).[27] Thus, opaque offers are exogenous search cost shifters, that do not affect v and ϵ and that can thus be used to test whether position affects search costs.

Prediction 3 (Case 3). *If position affects only search costs, the display of an opaque offer increases search costs and decreases consumer search of the demoted option.*

To perform the test for Prediction 3, I run a regression of the probability of clicking on the fifth displayed hotel in the ranking as a function of its position, hotel, and query characteristics for the Random ranking. Search impressions with opaque offers will display the fifth hotel in position 6, while those without will display it in position 5. In the regression, I compare search impressions without opaque offers (hotels are displayed in consecutive order), and search impressions with the first opaque offer appearing in position 5.[28] In Table 5, I show the results. In column 1, I find that when the fifth displayed hotel is shown in position 6 rather than position 5 (because it is demoted by an opaque offer), it receives 2.26% fewer clicks. I find a similar result after controlling for hotel and query characteristics (second column), as well as after restricting attention to search impressions with clicks lower in the ranking (third column), and to the largest destination in the data to include hotel fixed effects (fourth column). These results allow me to state that rankings affect consumer choices through search costs, thereby *ruling in* the effect of position on search costs.

One concern with this result is that consumers who clicked less on the fifth displayed hotel when it was shown in position 6 actually clicked on the opaque offer. Even though I cannot completely alleviate this concern, because I do not observe clicks on the opaque offer, such behavior is unlikely to be predominant in the data, since it involves navigating to a different screen. Thus, if a click on the opaque offer occurs, the consumer is unlikely to return to search lower-ranked hotels. To support this claim, in the third column of Table 5, I show that even in the sample of search impressions where consumers click lower in the ranking than the position of the opaque offer (these are search impressions that are the least likely to contain a click on the opaque offer), a similar result holds as in the full sample. Therefore, I expect clicks on opaque

Table 5. Estimates of Click on the Fifth Displayed Hotel (OLS)

	Baseline model		Searches with click after position 4	Largest destination
	(1)	(2)	(3)	(4)
Position effect				
Position 6	-0.0226^{*}	-0.0260^{*}	-0.0308^{*}	-0.0248
	(0.0102)	(0.0103)	(0.0143)	(0.0620)
Hotel characteristics				
Price		-0.0002^{***}	-0.0002^{***}	-0.0002^{*}
		(0.0000)	(0.0000)	(0.0001)
Stars		0.0219^{***}	0.0310^{***}	
		(0.0018)	(0.0025)	
Review score		-0.0001	-0.0011	
		(0.0013)	(0.0019)	
Chain		0.0079^{**}	0.0119^{**}	
		(0.0026)	(0.0037)	
Location score		0.0064^{***}	0.0087^{***}	
		(0.0010)	(0.0014)	
Promotion		0.0108^{***}	0.0157^{***}	
		(0.0027)	(0.0038)	
Query characteristics	Yes	Yes	Yes	Yes
Destination FE	No	Yes	Yes	Yes
Hotel FE	No	No	No	Yes
Adjusted R^2	0.000	0.008	0.014	0.014
Observations	50,870	50,160	33,442	1,719

Notes. Search impressions with opaque offers will display the fifth hotel in position 6, whereas those without will display it in position 5. The largest destination in the data is labeled 8,192. Query characteristics included in all regressions: trip length, booking window, number of adults and children traveling, number of rooms, and Saturday-night dummy. Standard errors are in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

offers to be more common in search impressions without any clicks on the displayed hotels and thus to not detract from the main findings of Table 5.

Cases 4, 5, and 7: Position Affects v and Another Component. If position affects v and another component, the realized utility is still a function of position. In this case, the model produces the same prediction as in Case 1. Thus, the same evidence and argument as in Case 1 can be used to rule out Cases 4, 5, and 7.

Case 6: Position Affects (ϵ, c). Suppose position affects both ϵ and c, so that $\sigma(p)$ and $c(p)$. As before, assume $\sigma(p) > 0$ and $\sigma'(p) > 0$. As the analysis for Case 3 showed, position increases search costs, so $c(p) > 0$ and $c'(p) > 0$. When position affects both σ and c, the effect of position on reservation utilities is unknown: higher search costs decrease z, whereas higher product uncertainty increases z. Formally, differentiating the reservation utility, $z = v + \sigma m$, I obtain

$$\begin{aligned} z' &= \sigma' m + \sigma m' \\ &= \sigma' m + \sigma\left[-\frac{1}{1-\Phi(m)}\frac{c'\sigma - c\sigma'}{\sigma^2}\right] \\ &= \sigma' m - \frac{c'}{1-\Phi(m)} + (\lambda(m) - m)\sigma' \\ &= \sigma'\lambda(m) - \frac{c'}{1-\Phi(m)} \\ &= \frac{\sigma'\phi(m) - c'}{1-\Phi(m)}, \end{aligned}$$

where the second equality follows from $m' = (\partial m/\partial x)\cdot(\partial x/\partial p)$, and the third equality follows from $x = c(p)/\sigma(p) = (1-\Phi(m))(\lambda(m) - m)$. Because $1-\Phi(m) > 0$, it follows that the $\text{sign}(z') = \text{sign}(\sigma'\phi(m) - c')$. Under the assumption that $\sigma' > 0$, the $\text{sign}(\sigma'\phi(m) - c') = \text{sign}(\phi(m) - c'/\sigma')$. Multiply and divide the second term by x to obtain

$$\begin{aligned} \text{sign}(z') &= \text{sign}\left(\phi(m) - \frac{c'}{\sigma'}\right) = \text{sign}\left(\phi(m) - \frac{c'/c}{\sigma'/\sigma}x\right) \\ &= \text{sign}(\phi(m) - e[\phi(m) - m(1-\Phi(m))]) \\ &= \text{sign}(\lambda(m)(1-e) + me), \end{aligned}$$

where $e = (c'/c)/(\sigma'/\sigma) > 0$, given the assumptions on search costs and product uncertainty. The last equality follows from the fact that $1 - \Phi(m) > 0$. What I have shown thus far is that $\text{sign}(z') = \text{sign}(\lambda(m)(1-e) + me)$. Thus, whether the reservation utility increases in position when both ϵ and c are a function of position depends on the values of m and e. For example, if $0 \leq$

$e < 1$ and $m \geq 0$, then $\text{sign}(z') = \text{sign}(\lambda(m)(1-e) + me) > \text{sign}(m(1-e) + me) \geq 0$, where the inequality follows from the fact that $\lambda(m) - m > 0$ (from results of Mills' ratio). However, more generally, both m and e are a function of search costs and the product uncertainty, so that the sign of z' depends on the functional-form assumption and values of c and σ.

To see how parameter values affect the sign of z', consider a typical parametrization of the model for common parameter estimates found in the literature. Suppose search costs are modeled as $c \propto \exp(\gamma p)$ (similar to Kim et al. 2010, 2017; Chen and Yao 2016; Ghose et al. 2012b) and product uncertainty is given by $\sigma(p) = p^n$.[29] Suppose $n \geq 1$; that is, product uncertainty is increasing at a nondecreasing rate, which captures the idea that product uncertainty is relatively small at the top of the ranking. Under this parametrization, $e = (\gamma p)/n$ and $x \propto \exp(\gamma p)/p^n$, which allows one to compute m from $m = B^{-1}(x)$. Previous literature as well as this paper (see Section 6.1) estimate a value of γ of approximately 0.01. For $n = (1, \ldots, 10)$ and $\gamma = 0.01 \cdot k$ for $k = (1, \ldots, 10)$, Figure 3 shows that for most parameter values that do not depart considerably from previous estimates ($k < 7$), reservation utilities increase in prices, so $z' > 0$.[30] Thus, as the following prediction describes, for the parametrization and the estimates found in the literature, the model would predict $z' > 0$.

Prediction 4 (Case 6)**.** *If position affects the expected utility from search and search costs, then for a common parametrization of the sequential search model and for estimates found in the literature, the reservation utility is increasing in position.*

The data reject this prediction. More precisely, the regression results in Table 4 presented to rule out Case 2 also rule out Case 6 for the above parametrization of the model and for common parameter estimates found in this paper and the literature. However, the test and the data patterns provided in this section do not rule out the possibility that position affects both ϵ and c for more general parameter values.

Table 6. Summary of Results

Case	Position affects	Prediction	Result
1	v	Purchase conditional on search is affected by position	Ruled out
2	ϵ	Reservation utility is increasing in position	Ruled out
3	c	Opaque offers increase search costs	Ruled in
4	v, ϵ	Purchase conditional on search is affected by position	Ruled out
5	v, c	Purchase conditional on search is affected by position	Ruled out
6	ϵ, c	For common parameter values, reservation utility is increasing in position	Ruled out in certain cases
7	v, ϵ, c	Purchase conditional on search is affected by position	Ruled out

To summarize, the analysis of all possible ways in which the effect of position can be modeled, produced the following results, condensed in Table 6. The model and the data rule out all cases, except Case 3 (position affects search costs) and Case 6 (position affects the expected utility from search and search costs). The fact that consumers click less on a demoted hotel when an opaque offer is present rules in the possibility that position affects search costs (Case 3). In addition, for a typical parametrization of the model and common parameter estimates found in the literature and in this paper, Case 6 is also ruled out. These findings provide support for the model in which position only affects search cost, which is the model used in this paper.

4.4. Empirical Specification

I now consider the empirical specification of the model. Let consumer i's utility from purchasing hotel $j \in \{1, \ldots, J\}$ be given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= x_{ij}\beta + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma_j^2), \end{aligned} \tag{8}$$

Figure 3. (Color online) The Sign of z' for Common Search Costs and Product Uncertainty Parameter Values

		K									
		1	2	3	4	5	6	7	8	9	10
n	1	0.30919986	0.29158312	0.27426459	0.2564307	0.23889738	0.22084076	-0.1009602	-0.4931323	-0.9401249	-1.4571442
	2	0.31536501	0.30399213	0.29298775	0.28155426	0.27048678	0.25898994	0.24746068	0.23629376	0.22469692	0.21306546
	3	0.31742006	0.30812847	0.29922881	0.28992878	0.28101658	0.27170633	0.26239112	0.25345756	0.24412933	0.23479468
	4	0.31844758	0.31019664	0.30234934	0.29411604	0.28628148	0.27806452	0.26985634	0.26203946	0.25384553	0.24565929
	5	0.3190641	0.31143754	0.30422165	0.29662839	0.28944042	0.28187944	0.27433547	0.2671886	0.25967526	0.25217805
	6	0.31947511	0.3122648	0.30546986	0.2983033	0.29154639	0.28442272	0.27732155	0.27062136	0.26356174	0.2565239
	7	0.31976869	0.31285571	0.30636144	0.29949966	0.29305064	0.28623935	0.27945447	0.27307333	0.2663378	0.25962807
	8	0.31998887	0.31329889	0.30703013	0.30039693	0.29417884	0.28760182	0.28105416	0.27491231	0.26841984	0.2619562
	9	0.32016012	0.31364358	0.30755022	0.3010948	0.29505632	0.28866152	0.28229837	0.27634263	0.27003921	0.26376697
	10	0.32029713	0.31391934	0.30796629	0.30165311	0.29575831	0.28950928	0.28329373	0.27748688	0.27133471	0.26521558

where x_{ij} is a vector containing the hotel's number of stars, review score, location, price, a chain, and a promotion indicator. The systematic component in the outside option is modeled as a fixed effect, more precisely, $v_0 = 1 \cdot \beta_0$. Because the data contain 54,877 hotels, including brand fixed effects in the model is not feasible. However, the outside option fixed effect captures the mean brand effect. Consumer i observes characteristics v_{ij} for each hotel j displayed in a search impression for free. To discover ϵ_{ij}, the consumer has to pay a search cost for clicking, which I model as

$$c_{ij}(p_{ij}) = \exp(k + \gamma p_{ij}), \tag{9}$$

where k gives the mean level of search costs and p_{ij} is the position of hotel j in the ranking at the time of the consumer's search query. The exponential function assumption of search costs is consistent with prior literature (Kim et al. 2010, 2017; Ghose et al. 2012b; Chen and Yao 2016) and it ensures that search costs are positive. I assume the position of the hotel affects search costs, as I found in the previous section.

Why is this model, based on Weitzman's (1979) search rules, preferable to other possible approaches to study consumer search for hotels on Expedia? In what follows, I discuss two modeling assumptions and argue in their favor. First, the model assumes consumers observe the list-page information of all hotels displayed at no cost and decide which ones to click and whether to purchase based on their reservation utility and utility. This modeling assumption seems inconsistent with consumer behavior: on one hand, if consumers know the list-page information, then why would position affect choices before search, while on the other hand, if position matters and the consumer has not observed a product because of its rank on the list, how would they know the list-page information of this option? This inconsistency thus represents a limitation of the current paper. The rationale behind this modeling assumption stems from a simplification of the modeled search process that allows researchers to use Weitzman's (1979) optimal search model in an empirical context. Indeed, in reality, observing list-page information involves a cost, which is associated with scrolling through the list until the hotel is reached, that is, a "scrolling cost." Thus, an alternative model could be one that includes consumers' preclick decision to observe the list-page information. Such a model would require solving two dynamic programming problems (computing two reservation utilities): one to decide how many hotels to collect list-page information from and another to decide when to stop searching for hotel-page information and what to purchase.[31] However, a way to simplify this model is to realize that consumers can only click on hotels for which they observed the list-page information. Thus, when they click, consumers must pay both a clicking and a scrolling cost. In the absence of data on consumers' scrolling behavior, my model approximates both of these costs by defining the search cost as having two components: mean search costs and product specific search costs (i.e., position effect). In this case, the mean level of search costs can be interpreted as the clicking cost and the position effect as the scrolling cost.

Second, the model assumes consumers search in order of reservation utilities as captured by Weitzman's (1979) search rules. Another variation of this model could be one where the order in which consumers search is exogenously given by the ranking. However, the data reject this approach. This alternative model suggests a consumer's nth click would occur in position n. Instead, in the data, the difference between clicks in search impressions with at least two clicks is on average 12 positions (median is 10). This suggests consumers' click order is not only determined by the ranking but also by observed characteristics, which the present model accounts for by modeling click order as being determined by reservation utilities.

In this section, I outlined a sequential search model that describes consumer choices from an ordered list. I also provided evidence to support the assumption of position only affecting search costs. Finally, I described the empirical specification of the model. Section 5 proceeds with the estimation of the model.

5. Estimation and Identification of the Model

5.1. Estimation

The model contains three types of parameters to be estimated: mean utility parameters (β, β_0), mean search cost (k), and the hotel-specific search cost or the position effect (γ). To see how the parameters of the model are estimated, suppose the consumer searched s of the total J hotels displayed in a search impression and that she chose j (including the outside option). Denote by $R_i(n)$ the identity of the hotel with the nth-largest reservation utility for this consumer. Order these hotels by their reservation utilities and let $R_i = \{R_i(1), \ldots, R_i(s)\}$ denote the set of searched hotels and the order in which they were searched.[32]

In this setting, Weitzman's (1979) optimal search strategy translates into the following restrictions on the utility and search cost parameters of the consumer. The selection rule requires that, if the consumer makes an nth search, her reservation utility from that hotel exceed her reservation utility from all hotels searched next and all those not searched. Formally, it must be that

$$z_{iR_i(n)} \geq \max_{k=n+1}^{J} z_{iR_i(k)}, \quad \forall\, n \in \{1, \ldots, J-1\}, \tag{10}$$

otherwise, the consumer would have searched a hotel with a higher reservation utility instead.

The stopping rule imposes two restrictions on the utility and search cost parameters. First, if the consumer makes an nth search, then her reservation utility from that hotel must exceed her utility from all hotels searched so far, including the outside option. Otherwise, the consumer would have stopped searching. Formally,

$$z_{iR(n)} \geq \max_{k=0}^{n-1} u_{iR_i(k)}, \quad \forall\, n \in \{1, \ldots, s\}. \tag{11}$$

Second, all unsearched hotels must have a lower reservation utility than the maximum utility of the searched alternatives, including the outside option

$$z_{iR_i(m)} \leq \max_{k=0}^{s} u_{iR_i(k)}, \quad \forall\, m \in \{s+1, \ldots, J\}, \tag{12}$$

otherwise, the consumer should have continued searching.

Finally, consistent with the choice rule, if the consumer chooses j (including the outside option), her utility from this choice must exceed the utilities of all of the hotels searched and the outside option. Formally,

$$u_{ij} \geq \max_{k=0}^{s} u_{iR_i(k)}, \quad \forall\, j \in R_i \cup \{0\}. \tag{13}$$

If consumers search sequentially, they make search and purchase decisions jointly. Thus, the probability of observing a certain outcome in the data is characterized by the joint probability of Equations (10)–(13) holding. Unfortunately, the data set does not contain information on the order in which consumers click. Thus, even though consumers decide in which order to click, as captured by the selection rule in Equation (10), only observing clicks and not click order does not allow this decision to inform the parameter estimates.[33] However, in this data, most consumers only click once, whereas for the 7% of search impressions with at least two clicks, I assume consumers click first on hotels displayed closer to the top of the ranking.[34] This assumption is supported by data from WCAI where consumers' search order is observed and where the position of the hotel explains most of the click order (see Table 9 in Online Appendix D), allowing me to use Equations (11)–(13) in estimation. Alternatively, the estimation method proposed by Honka and Chintagunta (2016) could be used, which accounts for consumers' click order in the absence of such data.

The probability P_{ijR_i} that the consumer searches in the order R_i and chooses j (including the outside option) is given by[35]

$$\begin{aligned} P_{ijR_i} &= \Pr\left(z_{iR_i(n)} \geq \max_{k=0}^{n-1} u_{iR_i(k)} \cap z_{iR_i(m)} \right. \\ &\qquad \left. \leq \max_{k=0}^{s} u_{iR_i(k)} \cap u_{ij} \geq \max_{k=0}^{h} u_{iR_i(k)} \right) \\ &= \int I(cond)\phi(\epsilon)\, d\epsilon, \end{aligned} \tag{14}$$

where $I(cond)$ is an indicator for whether conditions (11)–(13) hold. The log-likelihood is

$$LL = \sum_i \sum_{R_i} \sum_j d_{ijR_i} \log P_{ijR_i}, \tag{15}$$

where $d_{ijR_i} = 1$ if the consumer chose search order R_i and option j (including the outside option). Because search and purchase decisions are made jointly, the integral in Equation (14) does not have a closed-form solution. Thus, I replace P_{ijR_i} with the simulated choice probability $\hat{P}_{ijR_i}$. This approach results in the following simulated log-likelihood:

$$SLL = \sum_i \sum_{R_i} \sum_j d_{ijR_i} \log \hat{P}_{ijR_i}. \tag{16}$$

The choice probability $\hat{P}_{ijR_i}$ can be simulated in a number of ways. A straightforward simulator is accept-reject (AR), which Manski and Lerman (1981) originally proposed for probit models. This simulator approximates P_{ijR_i} by the proportion of draws from the appropriate distribution that satisfy the conditions in (14). However, using the AR simulator in maximizing the SLL can be problematic for two reasons. First, any finite number of draws can result in a reject, so that $\hat{P}_{ijR_i}$ is zero, which is especially likely if the data contain very few choices, as is the case here. The second difficulty comes from the fact that the choice probabilities are not twice differentiable, so the simulated probabilities will not be smooth. Thus, finding a maximum by optimizing the SLL using first and second derivatives will not be effective. Even though this problem can be circumvented by using an approximation of the gradient to the SLL instead, in practice, AR is difficult to use (Train 2009). An alternative simulator can be obtained by replacing the indicator function in the AR simulator with a smooth and increasing function that has defined first and second derivatives. As McFadden (1989) suggests, I choose the logit function that satisfies these conditions and is convenient to use. This method is known as the logit-smoothed AR simulator. It has been used to estimate probit models, as well as to estimate search models (Honka 2014, Honka and Chintagunta 2016).

Simulating $\hat{P}_{ijR_i}$ using the logit-smoothed AR simulator involves the following steps:

1. Draw $d = \{1, \ldots, D\}$ samples of ϵ_{ij}^d for each consumer and each hotel.
2. Use ϵ_{ij}^d to form utility u_{ij}^d.
3. Use the relation $x_{ij}^d = B(m_{ij}^d)$ to compute m_{ij}^d and form reservation utilities z_{ij}^d.
4. Define the following expressions for each draw d:
   (a) $\nu_1^d = z_{iR_i(n)}^d - \max_{k=0}^{n-1} u_{iR_i(k)}^d$;
   (b) $\nu_2^d = \max_{k=0}^{h} u_{iR_i(k)}^d - z_{iR_i(m)}^d$;
   (c) $\nu_3^d = u_{ij}^d - \max_{k=0}^{h} u_{iR_i(k)}^d$.

5. Compute S^d for each draw d using the expressions above

$$S^d = \frac{1}{1+\sum_{n=1}^{3} e^{-v_n^d/\lambda}}, \tag{17}$$

where $\lambda > 0$ is a scaling parameter.

6. The average of S^d over D draws of the error terms gives the simulated choice probability

$$\hat{P}_{ijR_i} = \frac{1}{D}\sum_d S^d. \tag{18}$$

Little guidance is available for choosing the scaling parameter λ. As $\lambda \to 0$, the simulator approaches the AR simulator and is thus unbiased. So, the researcher should use a small enough λ, but not too small to reintroduce the numerical problems one faces when optimizing with a nonsmooth function. For this data set, I determine the appropriate scaling parameter using Monte Carlo simulations, which are described in Section 5.2.2.

5.2. Identification

5.2.1. Model Parameters. In this section, I demonstrate how utility and search cost parameters are identified. The parameter set includes mean utility parameters, a mean search cost, and a product-specific search cost. The product uncertainty σ_j for $j \in \{0, 1, \ldots, J\}$ is normalized to 1 for identification, as is common in search and choice models (see Kim et al. 2010, 2017; Chen and Yao 2016).

Mean utility parameters that vary by product (hotel characteristics and the outside good indicator) are identified from Weitzman's (1979) stopping and choice rules in Equations (11)–(13). More precisely, the correlation between product characteristics and the frequency with which products are clicked and purchased, identify the mean utility parameters.

Weitzman's (1979) stopping rules in Equations (11)–(12) impose an upper and a lower bound on the mean search cost, respectively, that must have made conducting a certain number of searches optimal for the consumer.[36] The stopping rules, however, only recover a range of search costs. The level of search costs (parameter k) is pinned down by the functional form and the distribution of the utility function through the optimal search relation in Equation (3).

The product-specific search cost (the position effect) is identified from discrepancies in search and purchase frequencies. For example, products that are searched frequently, but not purchased, have low search cost (are ranked higher) and low mean utility parameters. On the other hand, products that are searched infrequently but purchased, have a high search cost (are ranked lower).

Characteristics in the utility function that do not vary by product cannot be identified, because they shift both the reservation utility and the utility by the same amount. Other factors that may affect consumer search costs that do not vary across products, but only vary across consumers, such as the booking window or the trip length, cannot be recovered without data showing multiple search impressions per consumer. Therefore, I do not include either of these factors in the specification of the model. Similarly, to capture consumer heterogeneity requires observing multiple search impressions per consumer, information that is not available in the data set.

An obstacle to identifying model parameters is that the position of a hotel (in a curated ranking), as well as its price, may be endogenous. In what follows, I describe how the Random ranking eliminates the endogeneity bias in position and show evidence to alleviate price endogeneity.

Position Endogeneity. Consumers pay more attention to higher-ranked products. At the same time, ranking algorithms predict consumer choices and rank those products that are most likely to be clicked or purchased at the top.[37] This creates a simultaneity problem (i.e., position is endogenous) in the estimation of the causal effect of rankings.

Formally, to illustrate this simultaneity problem, let θ_{ij} denote consumer i's preference parameter (e.g., utility and search cost parameters) for hotel j, and x_{ij} denote the hotel characteristics revealed at the time of i's query (in the current application, this represents the list-page information for each hotel). The simultaneity problem arises because rankings and choices are codetermined, each affecting the other. More precisely, rankings affect consumer choices; that is, the position of hotel j in the ranking affects whether the consumer chooses (clicks or purchases) this hotel. One can model this using the conditional probability of choice, $\Pr(\text{choice}_{ij} \mid x_{ij}, \text{position}_j, \theta_{ij})$. If both x_{ij} and position_j are independent of θ_{ij}, one can estimate consumer preferences consistently from this conditional probability. However, the ranking algorithm predicts consumer choices and constructs the ranking based on these predictions. Thus, consumer preferences affect the construction of the ranking, making position_j not independent of θ_{ij}. To estimate θ_{ij} consistently in this case, one also needs to model the marginal probability $\Pr(\text{position}_j \mid x_{ij}, \theta_{ij})$, because this marginal probability is not independent of the conditional probability, and estimate θ_{ij} from the joint probability[38]

$$\Pr(\text{choice}_{ij} \mid \theta_{ij}) = \Pr(\text{choice}_{ij} \mid x_{ij}, \text{position}_j, \theta_{ij}) \cdot \Pr(\text{position}_j \mid x_{ij}, \theta_{ij}). \tag{19}$$

However, modeling the marginal probability is challenging, because the exact ranking algorithm used by the search intermediary is proprietary, and thus potentially large biases may arise from trying to specify it.

Simultaneity problems of this type are also found in papers modeling the effect of marketing-mix variables that are set as a function of consumer-response variables (e.g., Manchanda et al. 2004), as well as in papers dealing with endogeneity as a result of targeting promotions to more responsive consumers (e.g., Nair et al. 2017).

The Random ranking does not suffer from endogeneity and can be used to estimate the causal effect of position. More precisely, when the ranking is randomly generated, the marginal probability of observing j in a given position is independent of θ_{ij}. Thus, maximizing the conditional likelihood is sufficient for unbiased estimates, so estimating θ_{ij} from $\Pr(\text{choice}_{ij} \mid \theta_{ij}) = \Pr(\text{choice}_{ij} \mid x_{ij}, \text{position}_j, \theta_{ij})$ will recover the unbiased estimate of θ_{ij}, which is the approach taken in this paper.

An additional concern with estimating the causal effect of position may be because of the design of the field experiment run by Expedia. More precisely, in this experiment, consumers who see the Random ranking are not aware that the ranking is random. As a result, even consumers who see the Random ranking may expect hotels to be ranked according to Expedia's ranking. In this case, they may expect higher-ranked hotels to be of higher quality, or in terms of the sequential search model presented in Section 4, the position of the hotel may affect their expected utility from search. However, the data are inconsistent with a model in which position affects the expected utility from search, as I show in Section 4.3. In addition, Chen and Yao (2016) show consumers are generally unaware of the ranking algorithm used by search intermediaries, thus making it unlikely that they have different expectations when seeing different rankings. Thus, the current experimental design is consistent both with data patterns on consumer behavior presented in this paper, and in the literature. Equally as important, changing the current experimental design to tell consumers how the ranking is constructed (e.g., that hotels are randomly ranked), would threaten both the internal and the external validity of the results. More precisely, if consumers were made aware of the manipulation (that hotels were randomly ranked), they might respond in ways that would threaten the internal validity of the results. For example, they might leave Expedia without clicking on any hotel (because in a field experiment, consumers are not aware of the experiment) or they might click randomly. In this case, the estimated position effect would not recover the effect of the ordering, but rather consumers' guesses of the hypothesis being tested. Second, because Expedia continually works to improve the ranking and consumers are not notified that it is doing so, making consumers aware of the manipulation would also threaten the external validity of the results. In other words, the results of such an experiment would be less relevant to firms making improvements to their rankings without announcing consumers. Because consumers typically do not know the exact way in which the ranking is ordered, it is important to know how these different ways of presenting alternatives affect their search and purchase behaviors in as natural a context as possible (i.e., without affecting their information set). These considerations allow me to conclude that the current design of the experiment is valid and it will lead to correct estimates of the causal effect of position from the Random ranking.

In sum, I conclude that using the Random ranking to estimate the effect of position solves the simultaneity problem and recovers the causal effect of position. In addition, any further concerns about bias in the effect of position are alleviated by experimental design and data patterns supporting the effect of position on search costs.

Price Endogeneity. The price of the hotel may be endogenous for two reasons. First, an unobserved quality shock may affect both consumer choices and hotel prices. Second, consumer specific choice probabilities may affect what prices hotels set. The most common method to alleviate price endogeneity is to use instrumental variables. One possibility is using Hausman-style (1996) instruments to approximate marginal costs, such as the average price of the same hotel or of the same-star hotels across destinations. Neither instrument is available in this paper: the former is not available, because in the data, I do not observe the same hotel in different destinations, whereas the latter is not available, because information on destinations is anonymized, and thus the average price of same-star hotels across different destinations (e.g., cities on different continents) may not recover the marginal costs of the focal hotel. Other possible instruments are lagged prices of the same hotel, which may not be valid if the unobserved quality of the hotel is correlated over time. Finally, another set of instruments are the average price of other hotels for the same trip and the focal hotel's nonprice characteristics. These instruments have been used by Chen and Yao (2016) in the hotel industry, and by Hortaçsu and Syverson (2004) and Berry et al. (1995) in different settings. They capture the position in characteristics space of the focal hotel relative to others, assuming that hotels' characteristics are predetermined. However, the last assumption may not be tenable in my data.

Even though price instruments in the hotel industry are difficult to obtain, concerns about endogeneity may be partially alleviated by the observation that prices, set by the hotel's revenue-management system, are not set in response to individual consumers' preferences (Cross et al. 2009, Mauri 2012, Koulayev 2014). To this end, in Table 8 in Online Appendix D, I show that observable hotel, trip, and query characteristics explain most of the variation in prices, with the trip-date explaining the majority of it. From discussions with an employee at a large hotel chain and from previous literature, I learned that the remaining price variation may be due either to (i) different suppliers selling the particular hotel or to (ii) experimental price variation (Einav et al. 2015, Koulayev 2014). Because both of these explanations are not demand related, I conclude that, conditional on the query, the price variation observed is unlikely to be correlated with the utility error term and thus does not need an instrument. Similarly, De los Santos and Koulayev (2017) use a regression to predict prices and include this prediction along with the observed price in the model they estimate following a control function approach. They find that this does not produce significantly different results (see Table 5 in their paper). Therefore, I choose to include only the observed price in my model.

5.2.2. Monte Carlo Simulation. In this section, I describe simulation results to show that simulated maximum likelihood using the logit-smoothed AR simulator can be used to recover utility and search cost parameters in this model.[39] To this end, I generate a data set of 15,000 consumers, each searching among five firms. Hotel characteristics (quality and price) are assumed to be drawn from a normal distribution with mean and standard deviation equal to those found in the data (see Table 1). The true values of the parameters are assumed to be consistent with those from a preliminary estimation of the model. For estimation, I use 50 draws from the distribution of the utility error terms for each consumer and hotel combination and I repeat the estimation 100 times. I have performed simulations with (inverse) scaling factor $1/\lambda$ ranging from 1 to 7 and found that $1/\lambda = 3$ works best in recovering the true parameters in the simulation sample, which motivated me to use the same scaling parameter in estimation.[40] The simulation results are given in Table 7. The first column shows the true parameters and the second column shows the estimated parameters. I find that this method works well in recovering the parameters of interest.

Having introduced the model and discussed its estimation and identification, in Section 6, I present estimation results using the Random ranking data set.

Table 7. Monte Carlo Simulation Results

	True values	Estimated values
Utility (u)		
Price	−1.5	-1.6666^{***}
		(0.0279)
Quality	1	1.0414^{***}
		(0.0133)
Outside option	1	0.7367^{***}
		(0.0656)
Search cost (c)		
Position	1	1.0042^{***}
		(0.0027)
Constant	−5.5	-5.2939^{***}
		(0.0015)
Log-likelihood		−30,140
Observations		75,000

Note. Standard errors are in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

6. Estimation and Counterfactual Results

6.1. Estimation Results

To recover the utility and the search cost parameters, I estimate the model using simulated maximum likelihood with the logit-smoothed AR simulator. I use 50 draws for each consumer-hotel combination of utility error terms, a scaling factor of $\lambda = 1/3$, and I repeat the estimation 50 times. To make the estimation feasible, I focus on the four largest destinations in the data.[41] This approach controls for potential unobserved destination-specific heterogeneity (compared to pooling observations from different destinations together), and shows that the results are not limited to a particular destination. Furthermore, I restrict attention to search impressions with opaque offers that have similar length (number of hotels displayed),[42] to ensure position effects are estimated consistently. Finally, I use the Random ranking data to eliminate the endogeneity bias in the ranking. The resulting estimation sample has a total of 83,504 observations and 2,510 search impressions. Each search impression has at least one click and there are 138 (5.5%) transactions.

Table 8 shows the main estimation results. In panel A, I show the estimated coefficients, whereas in panel B, I derive the magnitude of the position effect (the focus of this paper). In general, utility and search cost estimates are economically meaningful and significant. For example, search costs are higher for clicks occurring lower in the ranking (higher position) and consumers derive higher utility from cheaper hotels, with more stars and those that are running a promotion. I find mixed evidence for the effect of the location score on choices (depending on the destination), as well as a negative effect of the chain dummy (insignificant for most destinations) and of review score (possibly because of nonlinearity of the effect).

Using the Random ranking data for estimation eliminates the endogeneity bias in the ranking and recovers the true position effect. To measure the position effect, I compute the dollar equivalent of a change in search costs resulting from an increase in position by one. In panel B, I find an average position effect of \$1.92, which ranges from \$0.55 to \$3.19 across the four destinations. As expected, these results are typically smaller than those found in the hotel industry literature, which does not have experimental variation in

Table 8. Estimation Results

Destination	1	2	3	4
		Panel A. Coefficients		
Utility (u)				
Price ($100)	−0.2867***	−0.2312***	−0.1206***	−0.1857***
	(0.0291)	(0.0285)	(0.0419)	(0.0327)
Stars	0.3753***	0.0879***	0.0983***	0.1943***
	(0.0139)	(0.0284)	(0.0276)	(0.0235)
Review score	−0.2465***	−0.0452**	−0.0486*	−0.0808**
	(0.0336)	(0.0223)	(0.0256)	(0.0291)
Chain	−0.0089	−0.0097	−0.1810***	−0.0276
	(0.0353)	(0.0468)	(0.0562)	(0.0567)
Location score	−0.1477***	0.1022***	−0.0040	0.0438***
	(0.0301)	(0.0159)	(0.0254)	(0.0161)
Promotion	0.1584***	0.0654	0.0278	0.0226
	(0.0312)	(0.0452)	(0.0622)	(0.0549)
Outside option	0.0718	0.6244***	0.2378	0.4159***
	(0.0473)	(0.1064)	(0.0936)	(0.1138)
Search cost (c)				
Position	0.0044***	0.0185***	0.0121***	0.0109***
	(0.0017)	(0.0027)	(0.0027)	(0.0029)
Constant	−1.0305***	−1.4404***	−1.1467***	−1.0546***
	(0.0034)	(0.0028)	(0.0039)	(0.0046)
Log-likelihood	−3,546	−2,058	−1,346	−1,495
Observations	35,386	19,881	13,292	14,945
		Panel B. Equivalent change $		
Position ($)	0.55	1.90	3.19	2.04

Notes. Focus on the four largest destinations in the data and restrict attention to search impressions with opaque offers that have similar length. Standard errors are in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

the ranking. For example, De los Santos and Koulayev (2017) find position effects ranging from $7.76 to $35.15, Ghose et al. (2012b) find a position effect of $6.24, and Koulayev (2014) finds position effects that range from $2.93 to $18.78. The exception is Chen and Yao (2016), who find a position effect of approximately 21¢. The difference likely comes from the authors' ability to estimate the model at the consumer level, thereby aggregating different clicks made by the same consumer across search impressions, leading to lower estimated search costs.

As explained in Section 3.2.2, I expect the fact that most search impressions contain only one click and that different searches made by the same consumer cannot be linked will result in large search cost estimates. Indeed, I find baseline search costs exceeding $102. In the literature, the ability to link searches made by the same consumer significantly reduces search cost estimates, because this increases the number of clicks made by the unit of observation. For example, with an average number of clicks by consumer of 2.30 (rather than 1.12 by search impression in my case), Chen and Yao (2016) estimate a baseline level of search costs of $21.54, whereas De los Santos and Koulayev (2017) estimate search costs ranging from $8.35 to $55.23. As evidence, when estimating the model on the 7% of search impressions with at least two clicks, I find that mean search costs decrease by as much as 83% (see Online Appendix E), making this result comparable to search cost estimates in the literature. Note that in the latter estimation, the sample is considerably reduced, making many coefficient estimates, including the position effect, insignificant. However, as Section 5.2 shows, the effect of position is not identified from the number of clicks made (rather, it is identified from discrepancies in product search and purchase frequencies), so observing search impressions with mostly one click should not bias the position effect estimate (the coefficient of main interest).

In sum, eliminating the endogeneity bias in the ranking leads to a lower position effect, which I document. This result is consistent with the reduced form evidence presented in Section 3.3: rankings only affect consumer search, not their purchase decisions conditional on search, but in the absence of experimental variation, this effect is increased. Understanding the causal effect of rankings can help improve rankings, as I show in Section 6.2.

6.2. Counterfactual: Utility-Based Ranking

In this section, I investigate the welfare effects of a utility-based ranking through a counterfactual. To perform the counterfactual, I compare the utility-based ranking built on the estimates of the search model with simulated choices under Expedia's ranking. This approach involves two steps. First, I reorder hotels displayed under Expedia's ranking by their estimated expected utility prior to search (v_{ij}) to construct the utility-based ranking, and then simulate consumer choices and measure their welfare under the resulting ranking. Second, I simulate consumer choices using the list of hotels displayed under Expedia's ranking and measure consumers' welfare. Note that in comparing the utility-based ranking with simulated choices under Expedia's ranking (instead of actual choices), I avoid the bias of a direct comparison of the two rankings (Expedia and Random rankings), as discussed in Section 3.2.3. An alternative to this approach would be to approximate Expedia's ranking using data on which hotels it prioritizes and to simulate consumer choices and measure their welfare under the resulting ranking. However, this may introduce a bias, because approximating Expedia's ranking algorithm is challenging.

Consumer welfare from a ranking is defined as consumers' utility of the purchases made under that ranking, net of their total search costs from clicking. To obtain consumers' choices under a ranking, I simulate their click and purchase decisions using the utility and search cost parameters estimated from the Random ranking in Section 6.1 and Weitzman's (1979) optimal search rules. To integrate over the unobserved component in consumers' utility function, I repeat the simulation 50 times and I report the average result. To further minimize the impact of the unobserved component, I restrict attention to consumers who purchase because their utility is less dependent on it.

In this paper, a utility-based ranking coincides with a ranking based on consumers' reservation utility. The reservation utility of a hotel equals its expected utility prior to search plus a function of search costs (see Equation (4)). Hence, the hotel with the highest expected utility prior to search assigned the top position in the ranking, also has the highest reservation utility. As a result, under this utility-based ranking, consumers find it optimal to search in order of position.

The counterfactual results represent short-run effects of changing the structural parameters. Note that the counterfactual involves changing the order in which hotels are displayed without notifying consumers of the change. I adopt this approach for two reasons. First, this approach ensures that the results have external validity: when Expedia improves its ranking algorithm and thus the list of hotels shown, consumers are not made aware of this change. Therefore, my results are more relevant for firms that make changes to their ranking without announcing them. Second, notifying consumers of a change in the ranking could have affected their expected utility from search (ϵ_{ij}). However, the data patterns in Section 4.3, show that for the parameters estimated in this paper and in the literature, the position of a hotel affects only search costs (not expected utility prior to or from search), and thus these data patterns are consistent with the approach taken here of not notifying consumers of a change in the ranking.[43]

My results can be found in Table 9. I find that the utility-based ranking increases consumer welfare on average by $30.24 (23.02% of the purchase price), ranging from an increase of $9.05 (10.06% of the purchase price) to an increase of $53.58 (49.24% of the purchase price). A breakdown of this welfare shows that consumers benefit both from better matches (higher utility) and lower search costs. In particular, on average, 35% of the increase in welfare comes from lower search costs,[44] with the utility-based ranking decreasing by 4.64 positions, the average position in which a transaction happens.[45] This result confirms the importance of optimizing consumers' search decisions in improving rankings. However, the increase in consumer welfare is not due only to a decrease in search costs, as would be the case if consumers purchased the same hotel under the two rankings. Instead, some consumers purchase a more desirable hotel under the utility-based ranking that was inaccessible (was ranked lower) under Expedia's ranking. For these consumers, both utility and search costs improve. Therefore, I conclude that ranking hotels by their expected utility allows consumers to find better-matching products more quickly.

Table 9. Counterfactual Results

Destination	U-E 1	U-E 2	U-E 3	U-E 4
Consumer				
Change in consumer welfare ($)	9.05	31.30	53.58	27.04
(% transaction price)	(10.06)	(16.31)	(49.24)	(16.49)
Utility ($)	3.45	24.71	44.39	20.51
Total search costs ($)	−5.60	−6.59	−9.19	−6.53
Change in transaction position	−8.10	−3.24	−3.23	−3.98
Search intermediary				
% Change in transactions	2.40	6.00	4.73	4.22
% Change in revenue	5.71	−0.87	−12.94	−0.12

Note. U, Utility-based ranking; E, Expedia ranking.

At the same time, the utility-based ranking also benefits the search intermediary, with transactions increasing by at least 2.40%. This is expected, because contaminating the construction of the ranking with endogeneity bias by using estimates from consumer choices under a curated ranking can deteriorate the ranking. In addition, over industry measures, the utility-based ranking has the advantage of using estimates from a search model specifically designed to capture consumer behavior in an ordered environment, while also explaining their search and purchase decisions jointly. I also find that the effect on short-term revenues is uncertain: if consumers choose significantly cheaper hotels under the new ranking, even if transactions increase, revenues might fall. However, the effect on long-term revenues of a ranking that increases transactions may still be positive, because this encourages satisfied consumers to return for repeat business. As a result, I conclude that a utility-based ranking constructed with estimates from a sequential search model and with data without endogeneity bias in the ranking may provide a viable option for search intermediaries to improve their rankings.

In sum, I find that a utility-based ranking benefits both consumers and the search intermediary. This result further cements the value of estimating a sequential search model on the Random ranking data, because this provides clean estimates that help improve the ranking.

7. Conclusions and Future Research

In this paper, I study the causal effect of rankings on consumer search and purchase decisions. This is challenging because rankings are endogenous. To resolve this problem, I use the first data set with experimental variation in the ranking from a field experiment at Expedia. Using this data set, I make three contributions. First, I show that rankings affect what consumers search, but conditional on search, do not affect purchases. Second, I quantify the effect of rankings using a sequential search model following Weitzman (1979) and find position effects lower than previous estimates in the literature obtained without experimental variation. Also, I use model predictions, data patterns, and a feature of the data set (opaque offers) to show rankings lower search costs instead of affecting consumer expectations or utility. Finally, I use the model's estimates to measure the welfare effects of a utility-based ranking, and find that it improves matches and lowers consumer search costs, as well as benefits the search intermediary.

The current research could be improved in at least three directions, if the data limitations would be addressed. First, this data set does not allow comparison of the Expedia and the Random rankings in quantitative terms. If these data were available, one could quantify the endogeneity bias of the ranking, as well as evaluate different methods to eliminate this bias. This analysis would result in a method for resolving the endogeneity bias inherent in rankings without the need to conduct experiments. Second, the data set does not provide enough information to link different queries made by the same consumer. With this type of data, one could analyze the effect of rankings on consumer learning across search impressions, as well as provide a better measurement of the magnitude of consumers' search costs. Finally, consumers' click order is not observed in this data set. Because consumers choose the order in which they search, these types of data would allow one to account for click order in estimation and thus improve utility and search cost parameter estimates.

I also see two avenues for future research. First, as consumers transition to using mobile devices to access Internet content, intermediaries' rankings have to adjust to consumers' different behavior on these devices. For example, if the smaller screen size on mobile devices makes consumers focus even more on higher-ranked products than they do on a computer (by increasing their search costs), then the need for a relevant ranking will be even greater on mobile devices. Further research should focus on the relation between consumer search costs and the impact of the ranking. Another avenue for future research is to investigate the consequences of search intermediaries ranking only a subset of products, such as independent hotels in the case of online travel agents. Even though such rankings decrease the diversity of the products displayed, which apparently may hurt consumers, they also substantially speed up consumers' search (making these rankings especially relevant for mobile devices). Thus, exploring this trade-off can provide a new method to improve consumers' search experience.

Acknowledgments

The author is grateful to Ali Hortaçsu, Pradeep Chintagunta, Hugo Sonnenschein, and Richard Van Weelden. The author also thanks Paulo Albuquerque, Matthew Backus, Naiqing Gu, Elisabeth Honka, Liviu Iftode, Sergei Koulayev, Chris Nosko, Anita Rao, Stephan Seiler, and Bradley Shapiro. The author thanks the participants at the Searle Conference on Internet Search and Innovation (June 2015), the Workshop on Search and Switching Costs (May 2015), the IIOC conference (April 2015), and the University of Chicago IO and Marketing working groups for all their feedback. The author also thanks the two anonymous reviewers for comments that greatly improved the paper. The author thanks Expedia, Kaggle, and the Wharton Customer Analytics Initiative for providing the data. The usual disclaimer applies.

Endnotes

[1] All figures reported come from three sources: www.economist.com/news/business/21604598-market-booking-travel-online-rapidly-consolidating-sun-sea-and-surfing; www.forbes.com/sites/greatspeculations/2014/04/08/competitive-landscape-of-the-u-s-online-travel-market-is-transforming/; and www.wsj.com/articles/amazons-new-travel-service-enters-lucrative-online-travel-market-1429623993.

[2] In the case of Expedia, in the first quarter of 2013 (the relevant period for this analysis), 70% of its global revenues came from the sale of hotel rooms and most of its bookings were done under the agency model (54%). See Expedia's Earnings Release for the first quarter of 2013 for details: http://ir.expediainc.com/results.cfm.

[3] In the case of Expedia, ads may appear at the top of the ranking and in the last two positions. Details on the auction it uses for sponsored ads can be found at searchsolutions.expedia.com/how-it-works/.

[4] Search impressions with sponsored ads are included in the data, but are not flagged as such. To identify them, I use the fact that ads are more likely to appear in popular destinations with a large number of hotels.

[5] The data set I use resulted from an effort by Expedia to elicit the involvement of data miners in the improvement of its ranking algorithm. It was released through the International Conference on Data Mining (ICDM 2013) and Kaggle.com (an online platform facilitating participation of data miners in competitions posted by companies). The data are available at www.kaggle.com/c/expedia-personalized-sort/data.

[6] Online Appendix A contains the details about data cleaning.

[7] For more details on the WCAI data set, see Online Appendix B.

[8] In the WCAI data, I find that in 67% of search impressions, consumers only consider the first page of results.

[9] The WCAI companion data set shows few queries actually contain sorted/filtered results: only 32% of all search impressions and 34% consumers sort/filter.

[10] In the WCAI data set, I find more than 40% of consumers only search once.

[11] The largest country in the data is labeled 219.

[12] Information was retrieved in May 2015 from Alexa.com: www.alexa.com/siteinfo/expedia.com.

[13] Other data sets in the literature have similar properties. For example, Chen and Yao (2016), using the WCAI data, restrict their attention to consumers who have at least one transaction, thereby also reducing their data set to one that has at least one click and one transaction per unit of observation.

[14] Also, a higher fraction of converting search impressions appear under Expedia's ranking than under the Random ranking, possibly to hide its true conversion rate.

[15] I restrict attention to search impressions that do not include a hotel in positions 5, 11, 17, and 23, which are reserved for opaque offers by Expedia. Under the Random ranking, this represents 93% of search

impressions. For more details on opaque offers, see Section 4.3, as well as Online Appendix C.

[16] See Figures 4–6 in Online Appendix B for robustness checks, controlling for search impressions that more likely contain sponsored ads, that display few hotels, or that lead to a transaction. The same pattern as in Figure 1 holds.

[17] Note the conversion rate under Expedia's ranking is higher than that of the Random ranking.

[18] For ease of exposition, I focus on one consumer in Section 4.1. It is understood that components will have a subscript i for each consumer in what follows.

[19] As specified in Section 4.4, in this model, the product-page information is the source of consumer preference heterogeneity.

[20] This modeling approach is also found in papers estimating a sequential search model, such as Kim et al. (2010, 2017), De los Santos and Koulayev (2017), Chen and Yao (2016), and Ghose et al. (2012b). In addition, Athey and Ellison (2011) present a similar approach to modeling consumer behavior on a search site. More precisely, in their model, each firm has a two-dimensional type, part of which consumers discover for free by looking at the ad (equivalent to the list-page information in the current application), and part of which they need to click for (equivalent to the product-page information in the current application), which involves paying a search cost. The interested reader should refer to the click-weighting model presented in Section 6 of Athey and Ellison (2011).

[21] To be precise, an effect of position only on the realized value of ϵ_j after search would imply no effect of position on clicks, but an effect on purchases conditional on click. Both of these effects are ruled out by the reduced form evidence on the Random ranking presented in Table 2. If position affects the realized value of ϵ_j after search and at least one other component, then the analysis mirrors that in Cases 4, 5, or 7.

[22] Recall from above that it is sufficient to consider only the effect of position on the distribution of ϵ, which is what I focus on in this section.

[23] Note that if instead the model had $\epsilon \sim N(\mu, \sigma^2)$, the same argument as in Case 1 would rule out the effect of position on μ. To see this, observe that such a model would be equivalent (in expectation) to one in which $u = v + \mu + e$, where $e \sim N(0, \sigma^2)$. If $\mu(p)$, then $u(p)$, so the choice rule would predict that purchases conditional on click would be affected by position, which is ruled out by the data patterns in Section 3.3. Thus, the only way in which position affects the expected utility from search is if it affects σ.

[24] Note that $\sigma'(p) = 0$ corresponds to the case where position does not affect ϵ. Thus, I focus here on the case where σ is strictly increasing in position.

[25] For more details on the ranking algorithm, see Online Appendix B.

[26] See Online Appendix C for an illustration of opaque offers and the following resources for the information cited in this section on opaque inventory: https://en.wikipedia.org/wiki/Opaque_travel_inventory and https://www.tnooz.com/article/expedia-integrates-hotwire-distressed-inventory-in-hotel-booking-path/.

[27] See Online Appendix C for details on the three criteria (Expedia's revenue-management system, location, and availability) determining how opaque offers are displayed on Expedia, as well as for evidence that their display is unrelated to consumer queries in the data.

[28] The scarcity of the data prevents me from providing conclusive evidence from similar tests in the other positions reserved for opaque offers, namely 11, 17, and 23.

[29] I take $\sigma(p) = p^n$ as one possible model of the effect of position on product uncertainty that satisfies the constraint that $\sigma' > 0$. To my knowledge, papers in the literature studying the effect of position using a sequential search model assume product uncertainty is a known constant (fixed for identification), thus position does not affect σ. More precisely, Ghose et al. (2012b) and Chen and Yao (2016) consider the case where consumers incur a cost for exploring lower-ranked hotels, so rankings affect consumer search costs, whereas v contains observed hotel characteristics, and the variance of ϵ is a fixed constant independent of position. In addition, De los Santos and Koulayev (2017) consider the case where consumers click more at the top because they derive utility from higher-ranked hotels, so position affects v directly, without affecting search costs or the expected utility from search. Finally, more broadly, other papers estimating search models assume product uncertainty is constant, for example, Kim et al. (2010, 2017), Ghose et al. (2012b), Honka (2014), Honka and Chintagunta (2016), and Choi and Mela (2016).

[30] Also, for $0 \le k < 1$, the same result holds, because for e arbitrarily close to 0, $z' > 0$.

[31] Chan and Park (2015) describe such a model.

[32] In a slight abuse of notation, let the outside option be denoted by either $j = 0$ or $R_i(0)$.

[33] If data on consumers' click order were available, one way to incorporate the selection rule in the estimation of the model would be to model consumers' search costs as including an unobserved component.

[34] This assumption may lead to an overestimation of the position effect if the expected utility of the lower-ranked hotel was greater than that of the higher-ranked hotel. However, this scenario is unlikely because hotels are ranked randomly, leading to significant variation in the characteristics displayed.

[35] In Equation (14), I suppress the domain of possible values. It appears in Equations (11)–(13).

[36] The choice rule does not affect the search cost estimate.

[37] See more details about the ranking algorithm in Online Appendix B.

[38] I adapt this explanation from Nair et al. (2017).

[39] I thank Elisabeth Honka for her hints on running this simulation.

[40] Results with the remaining values of the scaling factor are similar and are available on request.

[41] The four destinations are 8,192, 4,562, 8,347, and 9,402, and are summarized in Table 10 in Online Appendix E.

[42] For each destination, I estimate the model on data with the two most frequently occurring search impression lengths. Because the number of hotels displayed varies, using the most frequent length by destination instead of the average across destinations has the advantage of capturing more representative data patterns.

[43] Note that if consumer beliefs about the ranking algorithm were affected (for example, if they knew Expedia switched to a utility-based ranking), the results presented in this section might no longer hold. This is a limitation of the current paper.

[44] For example, in the first destination, total search costs decrease by \$5.60, accounting for 62% of the increase in consumer welfare of \$9.05.

[45] Note that the change in total search costs is influenced both by the change in the position of a click, as well as by the change in the total number of clicks that consumers make.

References

Agarwal A, Hosanagar K, Smith M (2011) Location, location, location: An analysis of profitability of position in online advertising markets. *J. Marketing Res.* 48(6):1057–1073.

Ansari A, Mela C (2003) E-customization. *J. Marketing Res.* 40(2): 131–145.

Ansari A, Essegaier S, Kohli R (2000) Internet recommendation systems. *J. Marketing Res.* 37(3):363–375.

Athey S, Ellison G (2011) Position auctions with consumer search. *Quart. J. Econm.* 126(3):1213–1270.

Athey S, Imbens G (2015) Machine learning methods for estimating heterogeneous causal effects. Working paper, Stanford University, Palo Alto, CA, http://arxiv.org/pdf/1504.01132v1.pdf.

Baye M, De los Santos B, Wildenbeest M (2016) What's in a name? Measuring prominence, and its impact on organic traffic from search engines. *Inform. Econom. Policy* 34:44–57.

Berman R, Katona Z (2013) The role of search engine optimization in search marketing. *Marketing Sci.* 32(4):644–651.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large scale field experiment. *Econometrica* 83:155–174.

Chan T, Park Y (2015) Consumer search activities and the value of ad positions in sponsored search advertising. *Marketing Sci.* 34(4): 606–623.

Chen Y, Yao S (2016) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 34(4): 606–623.

Choi H, Mela C (2016) Online marketplace advertising. Working paper, Duke University, Durham, NC.

Cross R, Higbie J, Cross D (2009) Revenue management's renaissance: A rebirth of the art and science of profitable revenue generation. *Cornell Hospitality Quart.* 50(1):56–81.

De Corniere A, Taylor G (2014) Quality provision in the presence of a biased intermediary. Working paper, University of Oxford, Oxford, UK.

De los Santos B, Koulayev S (2017) Optimizing click-through in online rankings with endogenous search refinement. *Marketing Sci.* 36(4):542–564.

Einav L, Kuchler T, Levin J, Sundaresan N (2015) Assessing sale strategies in online markets using matched listings. *Amer. Econom. J.: Microecon.* 7(2):215–247.

Ghose A, Yang S (2009) An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Sci.* 55(10):1605–1622.

Ghose A, Goldfarb A, Han SP (2012c) How is the mobile Internet different? Search costs and local activities. *Inform. System Res.* 24(3):613–631.

Ghose A, Ipeirotis P, Li B (2012a) Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content. *Marketing Sci.* 31(3):492–520.

Ghose A, Ipeirotis P, Li B (2012b) Surviving social media overload: Predicting consumer footprints on product search engines. Working paper, New York University, New York.

Ghose A, Ipeirotis P, Li B (2014) Examining the impact of ranking on consumer behavior and search engine revenue. *Management Sci.* 60(7):1632–1654.

Hagiu A, Jullien B (2011) Why do intermediaries divert search? *RAND J. Econom.* 42(2):337–362.

Hausman J (1996) Valuation of new goods under perfect and imperfect competition. Bresnahan T, Gordon R, eds. *The Economics of New Goods, Studies in Income and Wealth* (National Bureau of Economic Research, Cambridge, MA), 207–248.

Honka E (2014) Quantifying search and switching costs in the U.S. auto insurance industry. *RAND J. Econom.* 45(4):847–884.

Honka E, Chintagunta P (2016) Simultaneous or sequential? Search strategies in the U.S. auto insurance industry. *Marketing Sci.* 36(1):21–42.

Hortaçsu A, Syverson C (2004) Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quart. J. Econom.* 119(2):403–456.

Jerath K, Ma L, Park Y, Srinivasan K (2011) A position paradox in sponsored search auctions. *Marketing Sci.* 30(4):612–627.

Jeziorski P, Moorthy S (2018) Advertiser prominence effects in search advertising. *Management Sci.* 64(3):1365–1383.

Jeziorski P, Segal I (2015) What makes them click: Empirical analysis of consumer demand for search advertising. *Amer. Econom. J.: Microeconomics* 7(3):24–53.

Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Kim JB, Albuquerque P, Bronnenberg BJ (2017) The probit choice model under sequential search with an application to online retailing. *Management Sci.* 63(11):3911–3929.

Koulayev S (2014) Search for differentiated products: Identification and estimation. *RAND J. Econom.* 45(3):553–575.

Manchanda P, Rossi P, Chintagunta P (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(4):467–478.

Manski C, Lerman S (1981) On the use of simulated frequencies to approximate choice probabilities. Manski C, McFadden D, eds. *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MA), 305–319.

Mauri AG (2012) *Hotel Revenue Management: Principles and Practices* (Pearson, Milan).

McFadden D (1989) A method of simulating moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5):995–1026.

Nair HS, Misra S, Hornbuckle WJ IV, Mishra R, Acharya A (2017) Big data and marketing analytics in gaming: Combining empirical models and field experimentation. *Marketing Sci.* 36(5):699–725.

Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Sci.* 34(3):388–407.

Rutz O, Bucklin R, Sonnier G (2012) A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *J. Marketing Res.* 49(3):306–319.

Train K (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).

Weitzman M (1979) Optimal search for the best alternative. *Econometrica* 47(3):641–654.

Yang S, Ghose A (2010) Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* 29(4):602–623.

Yao S, Mela C (2011) A dynamic model of sponsored search advertising. *Marketing Sci.* 30(3):447–468.